Received: 27 April 2020 | Accepted: 12 November 2024

DOI: 10.1111/jnp.12403

REGISTERED REPORT STAGE 2

The efficacy and feasibility of an immersive virtual reality game to train spatial attention orientation after stroke: A stage 2 report

Hanne Huygelier[1] | **Nora Tuts**[1] | **Karla Michiels**[2] | **Eline Note**[2] | **Fabienne Schillebeeckx**[2] | **Jos Tournoy**[3,4] | **Vero Vanden Abeele**[5] | **Raymond van Ee**[1,6,7] | **Céline R. Gillebert**[1]

[1]Department of Brain and Cognition, Leuven Brain Institute, KU Leuven, Leuven, Belgium

[2]UZ Leuven Campus Pellenberg, Leuven, Belgium

[3]Gerontology & Geriatrics, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[4]Department of Geriatric Medicine, UZ Leuven, Leuven, Belgium

[5]E-Media | Human-Computer Interaction, KU Leuven, Leuven, Belgium

[6]Donders Institute for Brain, Cognition and Behavior, Radboud University, Nijmegen, The Netherlands

[7]Philips Research, High Tech Campus, Eindhoven, The Netherlands

Correspondence
Céline R. Gillebert, Department of Brain and Cognition, Leuven Brain Institute, KU Leuven, Tiensestraat 102 box 3711, 3000 Leuven, Belgium.
Email: celine.gillebert@kuleuven.be

Funding information
Fonds Wetenschappelijk Onderzoek, Grant/Award Number: 1171717N, 1171719N, G078915N, G072517N, 1249923N, G0H7718N and G0B6622N; Herculesstichting, Grant/Award Number: AKUL/19/05; Methusalem funding KU Leuven, Grant/Award Number: METH/24/003

Abstract

Spatial neglect is a post-stroke attention deficit for which there is no evidence-based intervention. Immersive virtual reality (IVR) may increase treatment efficacy, as it allows to train spatial attention in a rich environment. This study evaluated the efficacy and feasibility of an IVR patient-tailored training (HEMIRehApp). Using a cross-over design, an active (spatially biased) and placebo (spatially unbiased) IVR intervention were compared. We aimed to recruit 8 per-protocol left-sided neglect patients. The primary outcome was response times on the Posner cueing task. To evaluate feasibility, we documented the number of recruited patients, cybersickness and patients' experience with HEMIRehApp. After 2 years of recruitment, we were able to enrol 6 patients, of whom 2 completed the full protocol. The target sample size was not feasible due to a lower than expected prevalence of left-sided neglect and a higher than expected drop-out rate. The planned group-level analysis was therefore replaced by a single-case analysis. The results in the 2 per-protocol cases suggest a superior effect of spatially biased IVR training than unbiased IVR training inside IVR. IVR training was feasible as all 6 enrolled patients were able to complete 10 IVR training sessions, but the cross-over protocol itself was unfeasible. While the low sample size prevented us from conclusively evaluating the efficacy of HEMIRehApp, our preliminary single-case results suggest that neglect patients were able to improve attentional orientation towards eccentric target locations in IVR. Follow-up studies are needed to further validate these findings.

KEYWORDS
attention, hemispatial neglect, immersive, rehabilitation, virtual reality

INTRODUCTION

Hemispatial neglect is a post-stroke attention deficit characterized by a difficulty in responding to events on the contralesional side of space, which significantly restricts patients' functionality and is a negative prognostic factor for stroke rehabilitation outcome (Buxbaum et al., 2004). Hemispatial neglect encompasses spatial and non-spatial deficits (Corbetta & Shulman, 2011; Husain & Rorden, 2003; Robertson, 2001) and is a heterogeneous syndrome. Patients can show symptoms in different spatial reference frames, distances to the observer and different sensory domains (Milner & Harvey, 1994; Stoep et al., 2013). Neglect can affect orientation to visual information, visually imagined information (Buxbaum et al., 2004), or can affect the motor domain (Laplane & Degos, 1983). Although hemispatial neglect is a complex syndrome occurring after both left- and right-hemispheric stroke (Demeyere & Gillebert, 2019; Ten Brink, Verwer, et al., 2017), we will focus on hemispatial neglect that affects processing of visual information in *peri-personal space* following right-hemispheric stroke.

Cognitive rehabilitation, non-invasive brain stimulation and pharmacological interventions

It is estimated that 40% of neglect patients do not experience complete recovery (Demeyere & Gillebert, 2017; Demeyere, Gillebert, et al., 2015; Nijboer et al., 2013). Thus, there is a great need for effective interventions for neglect (Kerkhoff & Schenk, 2012; Luauté et al., 2006; Van Vleet & DeGutis, 2013). Given that there is no strong evidence yet for the effect of cognitive rehabilitation at the group level (Bowen et al., 2013), researchers explored the potential of non-invasive brain stimulation (Cazzoli et al., 2012; Salazar et al., 2018; van Lieshout et al., 2019), pharmacological interventions (Luvizutto et al., 2015; van der Kemp et al., 2017) or re-evaluated classic rehabilitation methods such as prism adaptation (Ten Brink, Visser-Meily, et al., 2017). However, clinical trials have not yielded sufficiently strong evidence for the therapeutic effects or feasibility of these interventions for them to be used as a standard treatment in clinical practice. Harvey (2019) highlighted a salient issue in neglect rehabilitation research, stating that it was uncommon to report the number of referred relative to successfully recruited patients in clinical trials evaluating non-invasive brain stimulation. These data are indeed crucial to evaluate the clinical contribution of therapies, as they provide insight into the percentage of patients that can be treated with the therapy. Thus, it is not surprising that there is little consensus among clinicians about the preferred treatment for hemispatial neglect (Chen et al., 2018).

Virtual reality as an interesting avenue for neglect rehabilitation

Interestingly, when clinicians were asked about their preferred rehabilitation method in a scenario not hampered by practical constraints, many clinicians expressed a preference for virtual reality rehabilitation (Chen et al., 2018; Kolodziej & Gillebert, 2018). Indeed, immersive virtual reality (iVR), using head-mounted displays (HMD), offers several opportunities for neglect rehabilitation. The immersive nature of IVR may increase treatment engagement (Tieri et al., 2018). Moreover, the HMD offers excellent control over stimulus presentation (Foerster et al., 2016), allows training spatial orientation in a 3D, dynamic environment (Rizzo et al., 2004) and allows correction for compensatory head movements. IVR can indeed provide a safe and positive experience for older adults (Huygelier, Schraepen, van Ee, et al., 2019). Combining the strengths of iVR with gaming features that may further enhance treatment

engagement (Burke et al., 2009) may result in effective and feasible neglect rehabilitation. Therefore, we developed an iVR game to rebalance spatial attention in neglect patients (Huygelier, Schraepen, et al., 2020).

Potential limitations for the use of virtual reality in rehabilitation

Although iVR is promising, the clinical utility of iVR remains uncertain to date. For instance, iVR has been notorious for inducing *cybersickness*. However, the newest generation of iVR meets the technological standards necessary to effectively mitigate cybersickness (Kourtesis et al., 2019) and several studies reported little cybersickness using the latest generation of iVR technology in various older populations (Appel et al., 2020; Huygelier, Schraepen, et al., 2020; Huygelier, Schraepen, van Ee, et al., 2019; Plechatá et al., 2019). On the other hand, cybersickness has a negative association with sense of presence in the virtual environment (Weech et al., 2019) and may also depend on the design of the iVR application (Davis et al., 2015; Porcino et al., 2017; Stanney & Hash, 1998) as well as on characteristics of the end-users (Arns & Cerney, 2005). It is thus important to monitor cybersickness for each specific iVR application and for each end-user group.

Furthermore, cybersickness has mostly been studied using questionnaires administered after the iVR experience. As symptoms such as fatigue may already be present before using iVR, this procedure cannot clarify whether cybersickness resulted from the iVR experience itself. Indeed, a previous study showed a consistent decline in cybersickness after using an iVR application in six stroke patients (Huygelier, Schraepen, et al., 2020).

Finally, the clinical utility of iVR depends on how iVR is experienced by end-users (Huygelier, Schraepen, van Ee, et al., 2019). Nevertheless, only a few studies investigated iVR user experience in older adults and stroke patients (Dermody et al., 2020; Huygelier, Schraepen, et al., 2020; Huygelier, Schraepen, van Ee, et al., 2019; Tuena et al., 2020). One study reported a good usability of a VR assessment using shutter glasses in stroke patients with hemispatial neglect (Fordell et al., 2011). Another study reported a good perceived usability for an exergame using a computer monitor in hemispatial neglect patients (Tobler-Ammann et al., 2017). However, these studies do not clarify whether head-mounted iVR will be positively experienced by neglect patients.

Design of clinical trials

Many clinical trials on neglect rehabilitation have reported weak or moderate-quality evidence in favour of a therapeutic effect (Bowen et al., 2013; Luauté et al., 2006; Salazar et al., 2018; van der Kemp et al., 2017). Indeed, designing a sound clinical trial in stroke rehabilitation poses some challenges.

A first challenge lies in measuring changes in neglect symptoms. It is common to use clinical pen-and-paper assessments to measure the effects of rehabilitation, but research has shown large variation in scores on these assessments from test to retest, potentially obscuring rehabilitation effects (Bailey et al., 2004; Machner et al., 2012). Moreover, performance on clinical pen-and-paper tests is often summarized in a way that does not differentiate non-spatial from spatial errors, making it unclear which behavioural aspects may be affected by treatment (Huygelier, Moore, et al., 2020).

A second challenge is to determine the therapy dose prior to conducting a clinical trial. When obtaining non-significant results, researchers often conclude that their therapy dose may have been too low to obtain a therapeutic effect (e.g., Sturm et al., 2013). A possible solution for this salient issue may be to measure symptoms multiple times. Such a longitudinal design can inform whether more therapy hours may have resulted in better treatment effects.

Another challenge is to determine the baseline condition. In neglect rehabilitation, an experimental treatment is often compared to usual care (e.g., physical, occupational therapy), rather than to a placebo treatment. However, what exactly constitutes "usual care" is typically underreported and heterogeneous

across hospital sites (Negrini et al., 2020). Moreover, clinical trials have often underreported information relevant to assessing the feasibility of therapies (Harvey, 2019). These aspects make it difficult to generalize findings from clinical trials to new clinical contexts.

The current study

In the current study, we aimed to evaluate the *efficacy* and the *feasibility* of a new iVR rehabilitation game for hemispatial neglect.

More specifically, our primary objective was to compare the effect of an active and placebo iVR rehabilitation game on neglect symptoms assessed outside the iVR environment. In the active iVR intervention, multisensory stimulation was more frequently presented in the contralesional than the ipsilesional visual field, while in the placebo iVR intervention multisensory stimulation was presented in the central visual field. We assessed neglect symptoms using a Posner task outside the iVR environment, given that pen-and-paper neglect tasks often lack test-retest stability (Bailey et al., 2004; Machner et al., 2012). We hypothesized that the difference in response times to left- versus right-sided targets on the Posner task would decrease more as a function of the active than placebo intervention. In addition, clinical pen-and-paper tasks, daily life functioning and a computerized cancellation task were used as secondary outcomes. Here again, we hypothesized that the difference in performance between the left- and right visual fields would decrease more as a function of the active than placebo intervention. The current study aimed to address the impact of our intervention at the cognitive function and activities level and not at the participation level.

Second, we assessed the relationship between therapy dose and symptom recovery. To this end, the Posner task was administered after 4, 8 and 10 h of therapy. We hypothesized that there would be a larger effect of intervention hours on neglect symptoms in a non-iVR environment in the active than placebo condition.

Third, we evaluated whether our iVR rehabilitation game can impact non-spatial attention. We hypothesized that patients would improve more in non-spatial attention as a result of the active than the placebo iVR intervention.

Fourth, we evaluated training effects inside the iVR environment, as patients may show training effects that are only present in the iVR environment. Neglect symptoms were therefore assessed inside the iVR environment using a visual discrimination task and head orientation. We hypothesized that patients would improve within and outside the iVR environment, but that the improvement in the active iVR condition would be larger within than outside the IVR environment.

Finally, we evaluated several aspects of the feasibility of our iVR rehabilitation. More specifically, we reported the number of successfully recruited relative to referred patients and reported the reasons for missing data and drop-out. We also assessed the impact of a first experience with the iVR rehabilitation game on user experience and cybersickness. In line with an earlier study (Huygelier, Schraepen, et al., 2020), we expected a positive user experience and less cybersickness after than before the iVR experience. Last, we monitored cybersickness and user experience throughout the whole intervention period and assessed their relationship with the likelihood of drop-out. Here, we hypothesized that cybersickness and user experience would remain stable across sessions and would not be related to drop-out.

METHOD

Patient recruitment

Patients were recruited from one rehabilitation centre in Flanders (University Hospital Leuven campus Pellenberg). Patients residing in the rehabilitation unit older than 18 years (no maximum age)

with a stroke confirmed by a radiologist and who or their legal representative can provide informed consent were included for screening. Patients were excluded after screening when they did not show signs of hemispatial neglect for the left side of space (see Procedure section for details), did not have a right-hemispheric stroke, were left-handed, when the expected discharge was in a period shorter than 10 weeks, they had a severe comorbid psychiatric disorder, premorbid diagnosis of a neurodegenerative disease, medical safety contra-indications for iVR (e.g., medical electric implants based on EU safety guidelines, trepanation and history of epileptic seizures), a severe visual or auditory impairment that cannot be corrected while using the iVR system or a severe motor impairment that precludes them from using the iVR system. If patients did not meet any of these exclusion criteria, they were invited to participate in the clinical trial. Patients or their legal representatives provided written informed consent, and all study procedures were approved by the ethical committee of the UZ Leuven/KU Leuven (S61410) and were in accordance with the Helsinki declaration. Our study protocol was preregistered at clinicaltrials.gov (NCT03458611). The pre-registered and peer-reviewed stage−1 protocol can be accessed at https://osf.io/b4xfg.

IVR rehabilitation and IVR assessment

An iVR application was developed in Unity3D for Oculus Rift CV1 (Oculus Rift | Oculus, n.d.). Responses were registered using the right Oculus Touch Controller, and head movements were logged. Detailed information and pilot data of the iVR game are reported in Huygelier et al. (2017) and Huygelier, Schraepen, et al. (2020), and the game design is illustrated in a video (available online at https://doi.org/10.6084/m9.figshare.6194591.v2). The design of the iVR game was iteratively optimized based on pilot studies with stroke patients and neurologically healthy individuals (Huygelier, Schraepen, et al., 2020).

Patients visited a vegetable garden, lake or forest and performed good deeds for their neighbours or friends in the game world (Figure 1). For instance, patients were instructed to catch ladybugs in the neighbour's vegetable garden. To finish each level, they had to do two variations of a visual discrimination task. In Task 1, a 1 Hz audio-visual looming (i.e., grows in size and sound intensity) semi-transparent disk was presented in 50% of the trials for 3 s (Figure 1b). Afterwards, a target was presented in front of the white disk for 3 s or until patients made a response. In Task 2, the white disk moved from the centre towards the left or right side of the visual field, and once the target appeared, the cue disappeared, and the target moved towards the floor (Figure 1c). On each trial, feedback was presented to indicate whether the patient's response was correct (i.e., green checkmark), incorrect (i.e., red cross) or whether the patient did not make a response (i.e., blue exclamation mark and sound). In 25% of correct trials, patients received a score. The score was scaled as a function of the patient's performance. For instance, if they accurately discriminated only 25% of targets, they received 4 points; if they accurately discriminated 100% of targets, they received 1 point. Each rehabilitation level was completed once patients had obtained sufficient points.

In the *active iVR condition*, target locations were presented more often in the contralesional than the ipsilesional visual field, using a patient-tailored design. To tailor the ratio of contra- to ipsilesional stimulation to each individual patient, we first assessed the patient's spatial attention distribution in the iVR environment. In this *iVR assessment*, patients performed the visual discrimination task in each of the scenes (i.e., vegetable garden, lake and forest). Target locations were uniformly distributed within an area subtending 30° in the left and in the right side of the visual field and 5° in the upper and lower visual field. Targets were presented for 3 s or until a response was made. There were no rewards, and the assessment was finished when a total of 225 trials (i.e., 75 trials per scene) were completed. Before each scene, patients performed 10 practice trials. A model was estimated on their responses, and this model was mirrored along the x-axis to obtain a biased target probability distribution (Figure 1a).

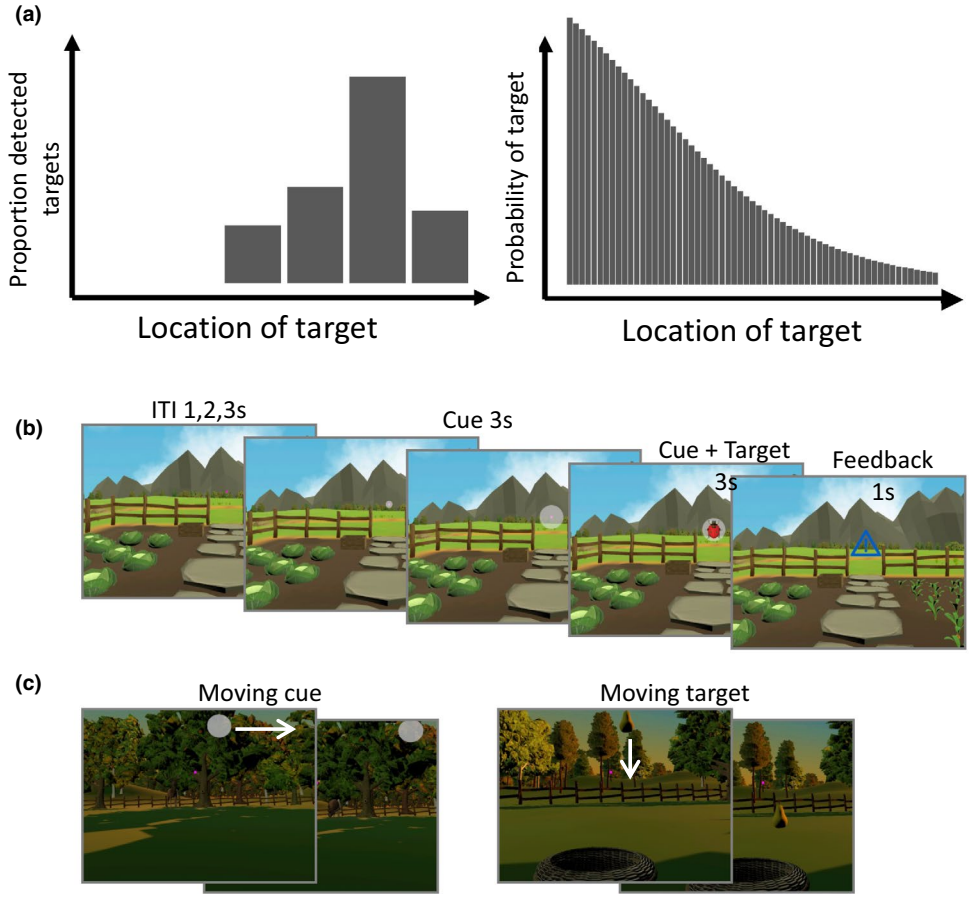

FIGURE 1 Illustration of patient-tailored design (a) and design of the tasks (b, c). If patients detect less targets for the left visual field during the assessment (panel a, left figure), then the probability that a target appears at those locations will be higher during rehabilitation (panel a, right figure). In Task 1 (panel b), a cue was presented for 3 s in 50% of trials. Then, the target and cue were presented for 3 s or until a response was provided. Afterwards, feedback was presented. In Task 2 (panel b), the cue moved towards a location after being presented in the centre of the visual field for 3 s and the target was presented without the cue and moved downward. Figure licensed under CC BY 4.0 by the authors. Retrieved from https://doi.org/10.6084/m9.figshare.12187710.v1.

This probability distribution was used to sample target locations, resulting in more contralesional than ipsilesional targets. For instance, a patient who participated in a pilot study with our game detected 32% of left-sided (<2°), 56% of right-sided (>2°) and 60% of targets in the centre of the visual field ([−2, 2°]) during the iVR assessment. During the active iVR rehabilitation, 61% of targets were located on the left (<2°), 32% were located on the right (>2°) and 7% in the centre of the visual field ([−2, 2°]). The target locations appeared at a maximum of 30° in the left and right sides of the visual field and 5° in the upper and lower visual field.

In the *placebo iVR condition*, target locations were sampled from a uniform distribution centred on 0° of the visual field with a horizontal angle of 1.5° in the left and right visual field and a vertical visual angle of 5° in the upper and lower visual field. All other game aspects were identical between the two conditions.

Materials for screening, primary and secondary outcomes

Semi-structured interview

A clinical psychologist evaluated the eligibility of patients to participate in the study (e.g., history of epilepsy, pacemaker and cochlear implant), collected basic demographic information (e.g., date of birth, gender, handedness and date of stroke) and obtained information about patients' medical history.

Questionnaires

To evaluate the feasibility of the treatment, we used several questionnaires. *Cybersickness* was measured with the Simulator Sickness Questionnaire (SSQ)[41] that was translated to Dutch by our research team. Each of the 16 SSQ items was rated on a scale with four levels representing no, mild, moderate or severe discomfort. The *User Experience scale* consisted of 23 items rated on a 5-point Likert scale going from totally disagree (1) to totally agree (5), with 3 as a neutral midpoint. Participants answered questions about the usability of the touch controllers, their sense of presence and their intrinsic motivation to play the IVR game. The motivation items were based on the intrinsic motivation inventory (McAuley et al., 1989), and the spatial presence items are translations of the International Test Commission Sense of Presence Inventory items (Lessiter et al., 2001). At the end of each game session, we asked participants whether they experienced any physical discomforts and which discomforts they experienced. We also asked them to rate their general game experience that day on a 5-point Likert scale.

To characterize patients' mood and fatigue at the start of the study, we administered two questionnaires in an interview format. We administered the *Dutch Hospital Anxiety Depression Scale*. We used the recommended cut-off score (≥8), which corresponds to an 80% sensitivity and specificity in detecting depression and anxiety (Bjelland et al., 2002). In addition, we administered the Dutch *Fatigue Severity Scale*, which has good reliability and validity in stroke patients (Nadarajah et al., 2017).

Neuropsychological pen-and-paper assessment

To screen for general cognitive impairments, the *Dutch Oxford Cognitive Screen (OCS-NL)* was administered (Demeyere, Riddoch, et al., 2015). Age-adjusted norms were used to interpret test scores (Huygelier, Schraepen, Demeyere, et al., 2019).

Several pen-and-paper cancellation tasks and a bisection task that are routinely used in clinical practice by clinical neuropsychologists and occupational therapists were used in screening or as secondary outcomes (Checketts et al., 2020; Evald et al., 2020). All tasks were administered on A4 paper in landscape orientation. The *OCS-NL hearts cancellation task* is a cancellation task with 50 full-outlined hearts and 150 hearts with a gap on the left or right side. The two parallel versions of the OCS-NL hearts cancellation task were administered intermittently across the assessment sessions (i.e., A-B-A-B-A). The *Random Shape Cancellation task* consists of 360 shapes (i.e., 60 targets, 300 distractors) randomly placed in an array of 24 cm by 19 cm. The *Star Cancellation test* (BIT, Halligan et al., 1991) was administered following the test manual instructions. To interpret whether there is a significant difference between right and left cancellations, we used a Bayesian contingency table test, similar to the approach of the Frequentist z-test of proportions (Huygelier, Moore, et al., 2020). We also administered the BIT *figure copy task* (Halligan et al., 1991).

The *McIntosh line bisection task* was administered (McIntosh, 2017; McIntosh et al., 2005). There are 4 line conditions (i.e., condition A: line from −4 cm to 4 cm, condition B: line from −8 to 4 cm, condition C: line from −4 to 8 and condition D: line from −8 to 8). Each line condition is presented 8 times on the page in a randomized order. The page was placed with the middle aligned to the patient's body midline. The patient was instructed to mark the middle of each line and tap the table in between each

response. Performance was summarized using the *endpoint weighting bias* (EWB). The cut-off scores based on healthy controls are equal to −.125 for right-sided neglect and .075 for left-sided neglect (McIntosh et al., 2017). Two parallel versions were administered intermittently.

Finally, we used the *Catherine Bergego scale* (CBS) (Azouvi et al., 2003; Ten Brink et al., 2013), which is a systematic observation scale frequently used by several health disciplines in clinical practice to measure how hemispatial neglect affects activities of daily living (Checketts et al., 2020). In addition, patients were asked to rate themselves on the scale items, and the difference between their own rating and the rating by the examiner was used as a measure of anosognosia (Grattan et al., 2018).

Computerized assessment

Neglect symptoms were measured using several computerized tasks. All computer tasks were administered on an LCD monitor with a resolution of 1920 by 1080 pixels. Patients were seated approximately 70 cm from the monitor. All code is written in Python 2.7 using Psychopy 1.90.3 (Peirce, 2007). Responses were registered with a standard keyboard or computer mouse.

Posner task

A Posner paradigm was used to measure the primary outcome. We chose exogenous cueing, as a meta-analysis indicated that neglect patients showed more pronounced spatial attention orientation differences in exogenous than in endogenous cueing paradigms (Losier & Klein, 2001). Three squares with a size of 1.5°, 2 located at 7° to the left and right of the fixation cross and 1 in the centre of the screen were presented. A cue (i.e., a colour change of a square) was presented for 100 ms. Subsequently, a target was presented 150 ms or 1100 ms after cue onset for 100 ms, in the left or right square (size of 1.4°) (Figure 2a). Cues and targets appeared on the left or right side of the screen with equal probability. The cue was valid (i.e., same side as target) in 40% of trials, invalid (i.e., opposite to target side) in 40% of trials or not followed by a target in 20% of trials (i.e., catch trials). Patients had to respond as quickly as possible when they saw the target in a 4 s time limit. They had to press the space bar with their right hand. After a response or when the time limit passed, the fixation cross disappeared for 1 s to indicate the end of a trial. Patients performed 10 practice trials in which feedback (i.e., "correct" or "incorrect") was shown for 1 s. There were 400 experimental trials that were presented in 4 blocks of 100 trials. The order of the trials was randomized.

Computerized cancellation task

A computerized cancellation task was administered as one of the secondary outcomes. Targets (i.e., full-outlined line drawings) and distractors (i.e., line drawings with upper or lower gaps) were presented in a grid with a width of 28 cm and height of 18 cm (Figure 2b). At a distance of 70 cm, this corresponds to stimuli being placed within a horizontal angle of 11° to the left and right sides and a vertical angle of 7.3° to the upper and lower sides of the visual field. The grid was divided into 15 equally spaced columns and 10 equally spaced rows. Stimuli were located in the grid at 15 horizontal locations, of which their position relative to the centre of the screen is: −13.01, −11.2, −9.33, −7.47, −5.6, −3.73, −1.87, 0, 1.87, 3.73, 5.6, 7.47, 9.33, 11.2 and 13.01 cm. Stimuli were located at 10 vertical locations: −8.1, −6.3, −4.5, −2.7, −.9, .9, 2.7, 4.5, 6.3 and 8.1 cm. The stimuli had a size of .9 by .9 cm. A random amount of jitter was added to make the search array disorganized with a maximum displacement of .45 cm.

For each trial, 50 targets and 100 distractors were presented. The target stimuli were spread randomly across the grid. For each set of 3 trials, the target was presented once in each cell. Each trial was presented for 4 min or until the patient indicated that he had finished the task by pressing the space bar. The patient was instructed to click on the targets using the left mouse button with their right hand. Once a target had been clicked, a blue line appeared on the target. A total of 12 trials were presented. One practice trial was presented in which 50 targets and 100 distractors were shown, and feedback was provided. When the patient clicked on a target stimulus, a green "V" sign appeared and a 400 Hz tone

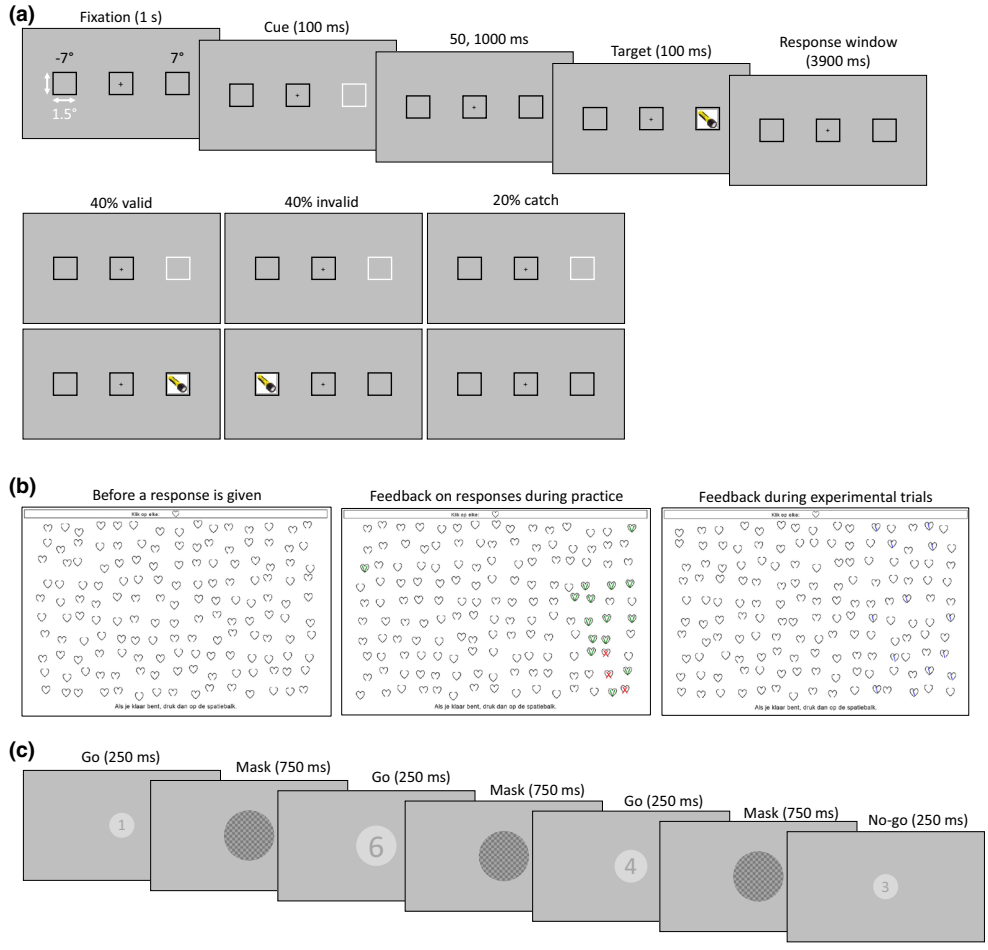

FIGURE 2 Procedure of the Posner task and conditions (a), examples of cancellation arrays of the computerized cancellation task (b) and procedure of the sustained attention to response task (c). In the SART, patients have to respond to all numbers except number "3".

was presented for 150 ms. When the patient clicked on a distractor stimulus, a red "X" sign appeared and a 200 Hz tone was presented for 150 ms.

Sustained attention to response task

To measure non-spatial attention, we used the sustained attention to response task (Robertson et al., 1997). Numbers from 1 to 9 were presented in a randomized order at a frequency of 1 Hz for 250 ms with 5 different sizes (Figure 2c). Patients were instructed to press the space bar every time a number appeared and to withhold their response when number "3" appeared. They had to do this task for a duration of 3.75 min, completing a total of 225 trials (i.e., 25 no-go trials and 200 go-trials). In addition, 27 practice trials were presented in which a short beep indicated errors.

Design

A placebo-controlled longitudinal study was used. Longitudinal designs are increasingly recognized as a more powerful approach relative to pre-post designs (National Research Council (US) Panel on Handling Missing Data in Clinical Trials, 2010), since they allow to measure symptom trajectories over

time. Given the lack of test-retest stability in commonly used measures of neglect (Bailey et al., 2004), this is especially important, as it may provide more reliable assessment of the symptom evolution. In addition, a longitudinal design allows to evaluate the dose-response relation.

We administered a placebo and an active version of the iVR game. The order of the placebo and active iVR game was counterbalanced between patients, creating two treatment groups (Figure 4a). We used a minimization algorithm to allocate patients to a treatment group. The first patient was allocated to group A or B using a random number generator. Then, patients were assigned to group A or B in a way that minimized the difference between groups in time since stroke.[1] This minimization approach has been shown to be more effective at balancing prognostic factors between treatment groups than stratified randomization for sample sizes of <100 patients (Kernan et al., 1999). As both groups only differ in the order of treatments, it is important to match the groups on patient characteristics that may influence recovery speed, such as time since stroke. Previous research has indeed established a fast recovery of neglect symptoms during the first weeks and a more stable recovery later on (Nijboer et al., 2013). Matching the two groups on potential moderators of response to treatment (e.g., age, anosognosia) is less important, as we will compare the active and placebo intervention conditions within patients and not between treatment groups.

Patients will not be explicitly informed about the placebo and active intervention. However, complete blinding of patients cannot be guaranteed, as patients may notice a difference between the interventions. For this reason, we asked patients whether they noticed a difference at the end of the follow-up session. The clinician who administered the pen-and-paper assessment and completed the behavioural observation scale was blinded to the treatment group. Blinding of the clinician was checked by letting them guess the treatment group the patient was allocated to after the follow-up session. Important to note is that the intervention was added onto care as usual (e.g., physical, occupational therapy). Given that care as usual varies between rehabilitation centres and patients (Negrini et al., 2020), we documented the types of therapy and number of therapy hours on a daily basis for each patient.

Procedure

First, adult stroke patients were invited to take part in a cognitive screening. When patients provided informed consent, three screening sessions were administered on consecutive days (Figure 3a). Based on this screening, the eligibility of patients to take part in the clinical trial was evaluated, and patient characteristics that may moderate treatment effects were recorded. If the proportion of cancelled targets on the left side of the computerized cancellation task was statistically significantly lower, according to a Bayesian contingency table test, than the proportion of cancelled targets on the right side, patients were considered left-sided neglect patients. The eligible patients were invited to take part in the clinical trial.

Patients were trained to use the iVR game. We administered a simulator sickness questionnaire (SSQ) before and after iVR exposure and administered a User Experience scale to gauge the patient's experience with the game (Figure 3b). In the same week, patients completed an iVR assessment, which was used to tailor the game to each patient. Then, patients completed 10 1-h game sessions in the active condition and 10 1-h game sessions in the placebo condition.

In addition, patients participated in eight weekly assessment sessions to measure the primary outcome (i.e., performance on a Posner task) and four secondary assessment sessions in which a battery of attention tests was administered. After completing the active and placebo intervention, there is a one-week follow-up assessment (Figure 3a). There were 8 parallel versions of the Posner task that only

[1]We calculated the hypothetical average time since stroke for each treatment group if the new patient would be included in either of the two groups. Then, we allocated the patient to that group that minimizes the difference. For instance, if patients in group A have an average time since stroke of 60 days and group B consists of patients with an average time since stroke of 30 days, a new patient with a time since stroke of 50 days is best allocated to treatment group B.

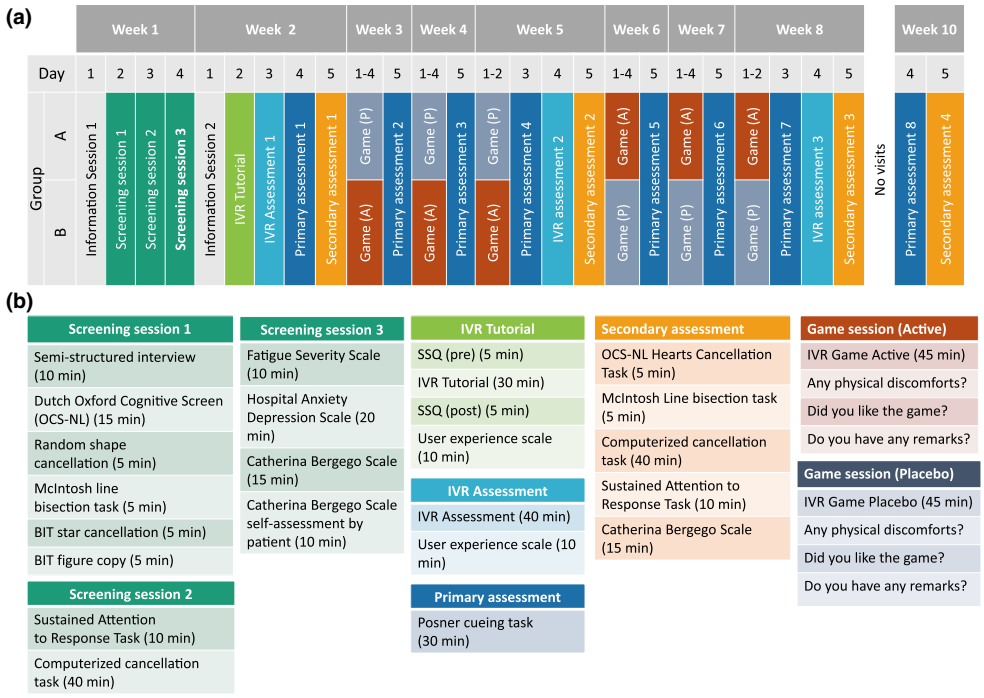

FIGURE 3 Trial Flowchart (a) and instruments per session (b). SSQ, simulator sickness questionnaire.

differed in the target shape (Figure 4a). There were five parallel versions of the computerized cancellation task that differed in the line drawings (Figure 4b). The order of the parallel versions was randomized between participants and matched between treatment groups A and B. The iVR assessment was re-administered when patients switched intervention conditions and after completing both intervention conditions (Figure 3a).

This study protocol involved a 1-h daily session. If a session cannot be completed due to practical constraints (e.g., technical problems), we shifted all sessions by 1 day. Through this procedure we aimed to minimize missing data.

Details of the data analysis are reported in Appendix A.

RESULTS

Patient recruitment and trial completion

Patients were recruited from the 5th of May 2021 until the 1st of August 2023 at rehabilitation hospital UZ Leuven Pellenberg. During this period, 188 patients were invited to participate in this study, of which 110 patients enrolled in neglect screening. Reasons for declining participation in the screening were not recorded. Seven out of the 110 patients did not complete screening. Only 11 patients showed signs of left-sided neglect on the computerized cancellation test. Of those patients, 6 patients enrolled in the training phase. The other patients were excluded due to epilepsy ($n = 3$), no right-hemispheric stroke ($n = 3$) and short hospitalization ($n = 3$) (Figure 5). Only 2 patients completed the entire cross-over protocol (i.e., per-protocol: VR060 and VR088). The other 4 patients (i.e., intention-to-treat patients) only completed a single phase of the training (active phase: VR034, VR076; placebo phase: VR080 and VR108), as they were released from the hospital before the study ended. Given that we did not reach the

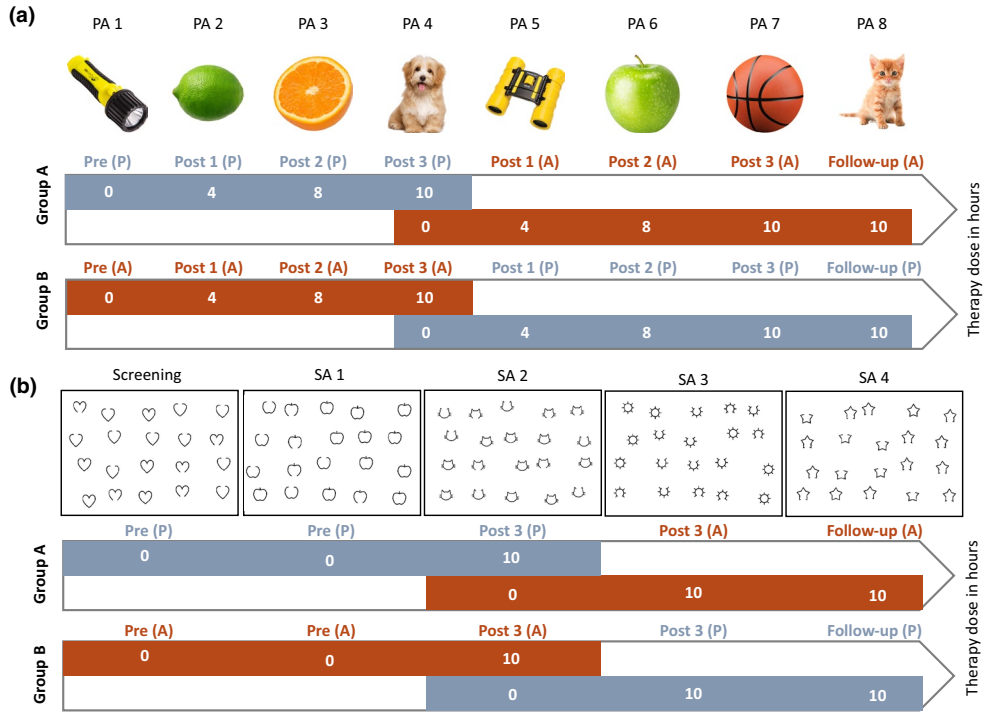

FIGURE 4 Parallel versions of the Posner task (a) and computerized cancellation task (b). An example of one sequence of the parallel versions is visualized. The sequence of versions will be randomized across participants and matched between groups A and B. A, active intervention (red colour); P, placebo intervention (blue-grey colour); PA, primary assessment; SA, secondary assessment. The planned therapy dose in hours is indicated per assessment session for the two intervention phases.

planned minimal per-protocol sample size for our group analysis, we adjusted our data-analysis plan to focus on single-case analyses of all cases (see Appendix B).

The age of the patients ranged from 39 to 73 years (Table 1), and patients were recruited 28 to 237 days post-stroke (Table 1). Lesions were delineated on CT scans (5 out of 6 cases) and 1 FLAIR scan manually in MRIcron. Scans were converted from native to MNI space using age-specific CT and MRI templates of the Matlab SPM clinical toolbox (Rorden et al., 2012). The lesion volume (Table 1) and lesion locations varied across patients (Figure 6). Lesion load in different brain parcels of the Yeo-Schaeffer atlas (Schaefer et al., 2018) is reported in Appendix C; Table C1. The 6 patients showed signs of a spatial bias across several screens for neglect (Table 2).

Delivery of treatment

The six patients played the iVR game for 3.7–6 h per condition (Table 3). The total number of completed training trials varied from 974 to 2998 trials across patients and conditions (Table 3). The 10 iVR game sessions were delivered in a period of 15–46 days (Table 3). During the period of iVR therapy, patients also received treatment as usual at varying dosages (Table 3). Treatment as usual comprised of physical therapy (62% of therapy time), occupational therapy (24%) and sessions with the speech therapist or neuropsychologist (14% of time). The median position of targets in the active iVR game ranged from −15.87° to −8.93° (° = horizontal angle relative to the head) across patients (Table 3). The median position of targets in the placebo iVR game ranged from −.17° to .03° across patients.

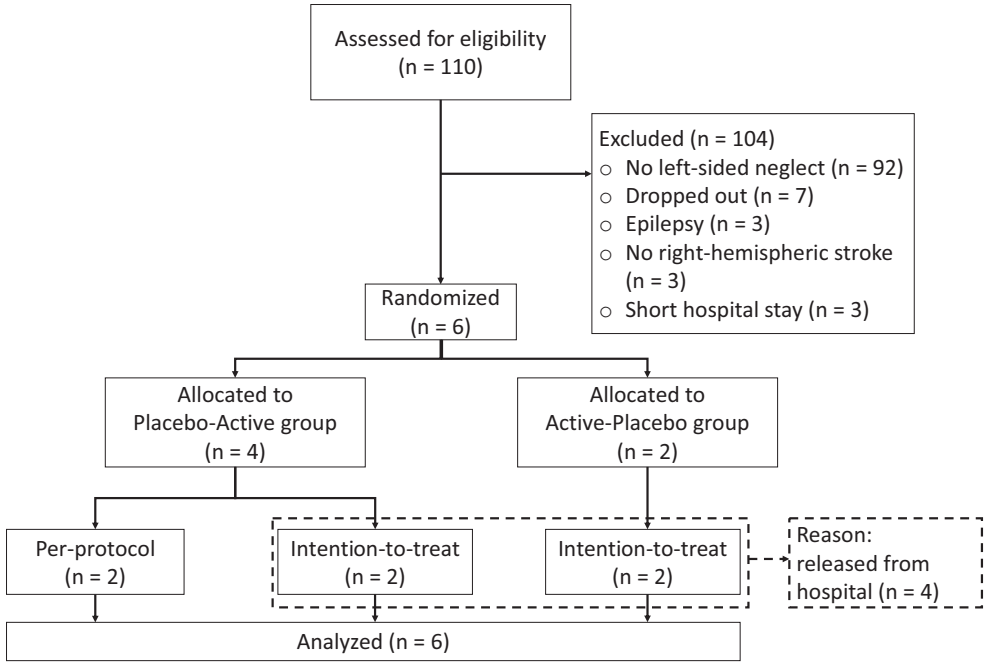

FIGURE 5 CONSORT flowchart of patient enrolment and reasons for exclusion and drop-out.

Feasibility of IVR training

At the end of each game session, patients were asked to indicate whether they experienced any physical discomforts and to rate their general game experience on a 5-point scale. Two patients (VR060 and VR080) never reported any physical discomforts. One patient (VR088) only reported physical discomforts on 1 session. The remaining three patients reported physical discomforts across multiple sessions, such as fatigue, headache, tired eyes or uncomfortable pressure from the VR headset (Table 4). Patient VR108 frequently complained of neck pain (70% of sessions). User experience was rated neutral to positive, except for VR108 who rated his game experience as negative on average (Table 4).

Training effects inside the IVR environment

Performance in the IVR assessment

We investigated whether attention towards left targets improved more after the active iVR training compared to the placebo iVR training in the iVR assessment. To this end, we compared performance as a function of target location between the baseline, post-placebo and post-active iVR assessments. We expected that performance for left-sided target locations would improve more after active than placebo training.

Performance for left-sided targets indeed improved more after active than placebo training for patient VR060, especially for target locations that were located further away from the head midline. That is, the probability to respond to a left-sided target at 30° was very low (.04, 95% CI = [.01, .08]) at baseline and remained low after placebo training (.04, 95% CI [.01, .09]). After active training, the probability increased to .32 (95% CI = [.17, .48]) (Figure 7). For VR088 a similar pattern emerged. That is, performance for targets located far from the head midline decreased from the baseline to placebo assessment and then improved after active training (Figure 10). The probability to respond to left-sided

TABLE 1 Patient descriptives.

Patient	Sex	Age (years)	Dominant hand	Time since stroke (days)	Lesion volume (cm^3)[c]	Barthel total[a]	FSS	HADS	Impaired OCS-NL tests[b]
Per-protocol patients									
VR060	M	69	R	28	49	10	6[d]	11[d]	Executive, praxis
VR088	F	39	R	77	46	6	2	4	None
Intention-to-treat patients									
VR034	M	66	L	237	106	/	4[d]	14[d]	None
VR076	F	73	R	64	361	1	3	8[d]	Verbal memory
VR080	F	53	R	70	327	/	6[d]	3	Reading, praxis
VR108	M	70	R	32	21	6	5[d]	15[d]	Executive, orientation

Abbreviations: Barthel, a measure of independence in activities of daily living scored on a 20-point scale (0 = fully dependent); FSS, Fatigue Severity Scale (an index of the severity of fatigue ranging from 0 (no fatigue) to 7); HADS, Hospital Anxiety and Depression Scale (depression subscale) ranging from 0 (no depression) to 21; OCS-NL, Dutch version of the Oxford Cognitive Screen.

[a]Barthel Index from patient's medical files.

[b]OCS-NL subtests other than visual field and hearts cancellation on which patients performed below age-adjusted normative cut-offs.

[c]Lesion volume was based on normalized lesion maps.

[d]Scores were below or above clinical published cut-offs.

targets at 30° was significantly higher after active (.36, 95% CI = [.23, .51]) than placebo training (.13, 95% CI = [.07, .21]).

Patient VR076 improved in target detection across all target locations after the active iVR training relative to the baseline assessment (Figure 7). Patient VR080 improved in target detection after placebo training, but not for targets located far away from the head midline. That is, at 30°, the probability to detect targets was low at baseline (.04, 95% CI = [.02, .08]) and after placebo training (.04, 95% CI = [.02, .09]). Patient VR034 and VR108 did not demonstrate clear signs of left-sided neglect at the iVR baseline assessment (Figure 7).

Head orientation in the IVR assessment

We additionally assessed whether patients oriented their head more contralesionally during the iVR assessment after the active and placebo iVR training compared to the baseline assessment.

For VR060, the median direction at the baseline assessment was oriented far to the right (Mdn = 56°, concentration[2] = .93°). At the post-placebo assessment, the median direction significantly changed from baseline and was now close to a straight ahead orientation (Mdn = .8°, concentration = .99°, Rao[3] = 5080.7, $p < .001$). At the post-active assessment, the median direction was rightward, but significantly less than at baseline (Mdn = 8.5°, concentration = .97°, Rao = 4305.8, p-value < .001) (Figure 8).

For VR088, the median direction at the baseline assessment was oriented far to the right (Mdn = 47.5°, concentration = .89°). After the placebo training, the median direction was even further to the right than the baseline assessment (Mdn = 58.8°, concentration = .91°, Rao = 3364.21, $p < .001$). After the active training, the median direction was leftward (Mdn = −11.7°, concentration = .90) and differed significantly from the baseline assessment (Rao = 4298.27, $p < .001$) (Figure 8).

For VR034, the median direction was far to the right at the baseline assessment (Mdn = 37.5°, concentration = .95°). At the post-active assessment, the median direction was even further to the right (Mdn = 51°, concentration = .96°, Rao = 3148.48, $p < .001$) (Figure 8).

For VR076, the median direction was far to the right at the baseline assessment (Mdn = 32.7°, concentration = .90°). At the post-active assessment, the median direction was similar to the baseline assessment (Mdn = 33.7°, concentration = .93°). The distributions did differ significantly (Rao = 3958.80, $p < .001$), with the global mode more towards the centre in the baseline assessment than the post-active assessment (Figure 8).

For VR080, the median direction at baseline assessment was far to the right (Mdn = 33.9°, concentration = .98°) and less oriented to the right at the post-placebo assessment (Mdn = 28.8°, concentration = .94°, Rao = 3426.25, $p < .001$) (Figure 8).

For VR108 the median direction was far rightward at the baseline assessment (Mdn = 50°, concentration = .97°). At the post-placebo assessment, the median direction was significantly less rightward (Mdn = 16.7°, concentration = .97°, Rao = 2964.05, $p < .001$) (Figure 8).

Training effects outside the IVR environment

To assess the efficacy of the active iVR training relative to the placebo iVR training, we assessed the effect of the two training conditions on several outcome measures outside the iVR environment.

[2]Concentration refers to the mean resultant length of the circular data, representing the spread of the data, ranging from 0 (maximal spread) to 1 (minimal spread).

[3]A non-parametric statistical test was conducted to compare the two circular head orientation distributions.

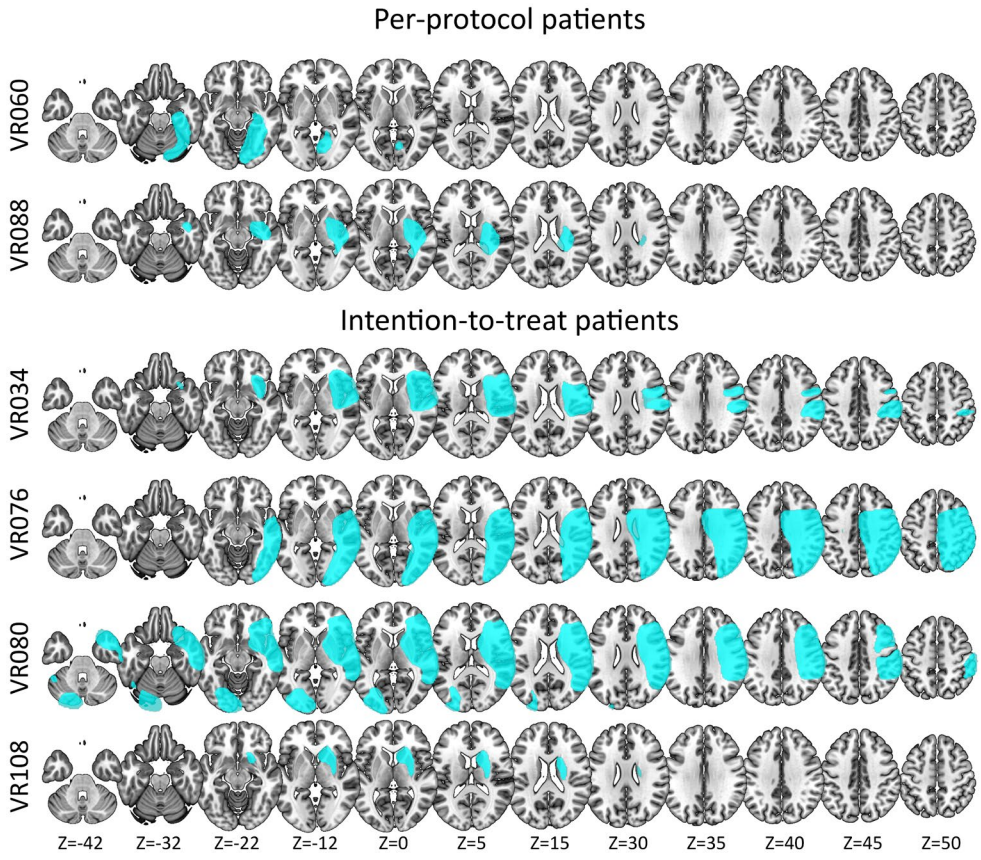

FIGURE 6 Lesion overlays of each patient on the MNI 152 template included in MRIcroGL depicted in neurological convention (left hemisphere on left side). Patient VR080 had an acute ischemic stroke in the right hemisphere but also an old lesion related to a resected meningioma in the left hemisphere.

Primary outcome measure: Reaction times on the Posner cueing test

The primary outcome measure was the reaction times for invalid cued left targets on the Posner test. We predicted a stronger decrease in response times for invalidly cued left targets during active training than placebo training. In addition, we predicted no increase in response times for invalidly cued right targets. Therapy dosage was operationalized as the number of trials that patients completed in the placebo and active therapy conditions (as these were not matched, see Table 3). Response times were modelled with a shifted lognormal model that can accurately capture the skew typical for response time distributions. To obtain a good model fit, the mean and standard deviation of the shifted lognormal model[4] were allowed to vary over time, and we assumed that response times were right-censored (Appendix B).

For VR060, there was evidence in favour of a weaker decrease in mu of the lognormal distribution of left-sided response times during the active than placebo therapy phase (.11, 95% CI = [.07, .15], PP[5](H+) = 1). There was also evidence for a weaker decrease in mu (.03, 95% CI = [.02, .04], PP(H+) = 1) and sigma (.05, 95% CI = [.04, .07], PP(H+) = 1) of the lognormal distribution of right-sided response

[4]The mu and sigma of the lognormal distribution do not have a one-to-one relation with the mean and standard deviation of the observed response times. Thus, the mean response time can be higher either by an increase in the mu or sigma of the lognormal distribution.
[5]PP = posterior probability of the one-sided hypothesis that the coefficient is positive (+) or negative (−).

TABLE 2 Performance on pen and paper screening tests for neglect and visual field deficits.

Test	Outcome	Max	Per-protocol patients		Intention-to-treat patients			
			VR060	VR088	VR034	VR076	VR080	VR108
OCS-NL visual field	Total	4	1	3	4	2	2	4
	R-L	2	**1**	**1**	0	**2**	**2**	0
OCS-NL hearts cancellation	Total	50	14	23	37	13	27	42
	R-L (ego)	20	**6**	−1	**4**	**7**	**5**	0
	R-L (allo)	50	**14**	2	0	**9**	**6**	0
Star cancellation	Total	54	31	47	51	37	19	54
	R-L	27	**11**	**5**	−1	**11**	**5**	0
Weintraub cancellation	Total	60	30	47	45	32	3	56
	R-L	30	**8**	**13**	**9**	**18**	1	**4**
McIntosh line bisection	EWS	1	.63	.34	.48	.49	.13	.85
	EWB	1	**.20**	**.45**	**.28**	**.28**	**.29**	**.23**
CBS (observer)[a]	Total	30	**6**	**18**	**5**	**10**	**12**	**12**
CBS (patient)	Total	30	**10**	**12**	**16**	**10**	**5**	**7**
Figure copy test	Total	3	**1**	**1**	3	**1**	**0**	3

Abbreviations: allo, allocentric neglect; CBS, Catherine Bergego Scale; Ego, egocentric neglect; Max, maximal score one can obtain on the test; OCS-NL, Dutch version of the Oxford Cognitive Screen; Scores in bold indicate neglect according to published cut-offs.

[a]CBS was administered and scored following the Kessler Foundation Neglect Assessment Procedure by the research team (Chen et al., 2015).

TABLE 3 Delivery of treatment and treatment dosage of placebo and active IVR training.

Patient	Training condition	Left targets (%)	Position targets (°Mdn, SD)	Total playtime (h)	Trials (*n*)	Duration (days)	TAU (h)[a]
Per-protocol patients							
VR060	Placebo	49	.03, .88	5.50	1790	15	5
	Active	70	−8.93, 14.06	5.96	2998	19	17
VR088	Placebo	50	.00, .89	5.25	1278	26	28
	Active	71	−10.44, 15	5.04	2686	15	28
Intention-to-treat patients							
VR034	Active	100	−15.87, 8.7	3.95	1964	15	28
VR076	Active	66	−10.00, 16.53	5.63	974	28	25
VR080	Placebo	55	−.17, 6.11[b]	3.68	1104	32	17
VR108	Placebo	51	−.01, .85	4.24	1230	46	17

Abbreviation: TAU, treatment as usual.

[a]Total number of hours of treatment as usual recorded across the training period.

[b]The spread of targets is larger for this patient as the VR game settings were accidentally set to the active condition for 1 session (encompassing 12% of all training trials). The data of this patient were still included in the analyses as the median position of all training trials was close to 0°.

times during the active than placebo therapy phase. The latter resulted in a reduced rate of change in the mean response times for the left and right visual field during the active therapy phase relative to the placebo phase, in contrast to what was predicted (Figure 9).

For VR088 there was evidence in favour of a weaker decrease in sigma of the lognormal distribution of response times for left-sided targets during the active than placebo therapy phase (.03, 95% CI [.01, .05], PP(H+) = .99). The latter results in a reduced rate of change in the mean response times on

TABLE 4 Physical discomforts and user experience on a 5-point scale across VR game sessions.

Patient	Physical discomforts (% sessions)	User experience M (SD)
Per-protocol patients		
VR060	0	3.2 (.5)
VR088	5	4.2 (.4)
Intention-to-treat patients		
VR034	30	3.6 (.7)
VR076	39	3.4 (1.0)
VR080	0	4.1 (1.7)
VR108	70	2.6 (1.2)

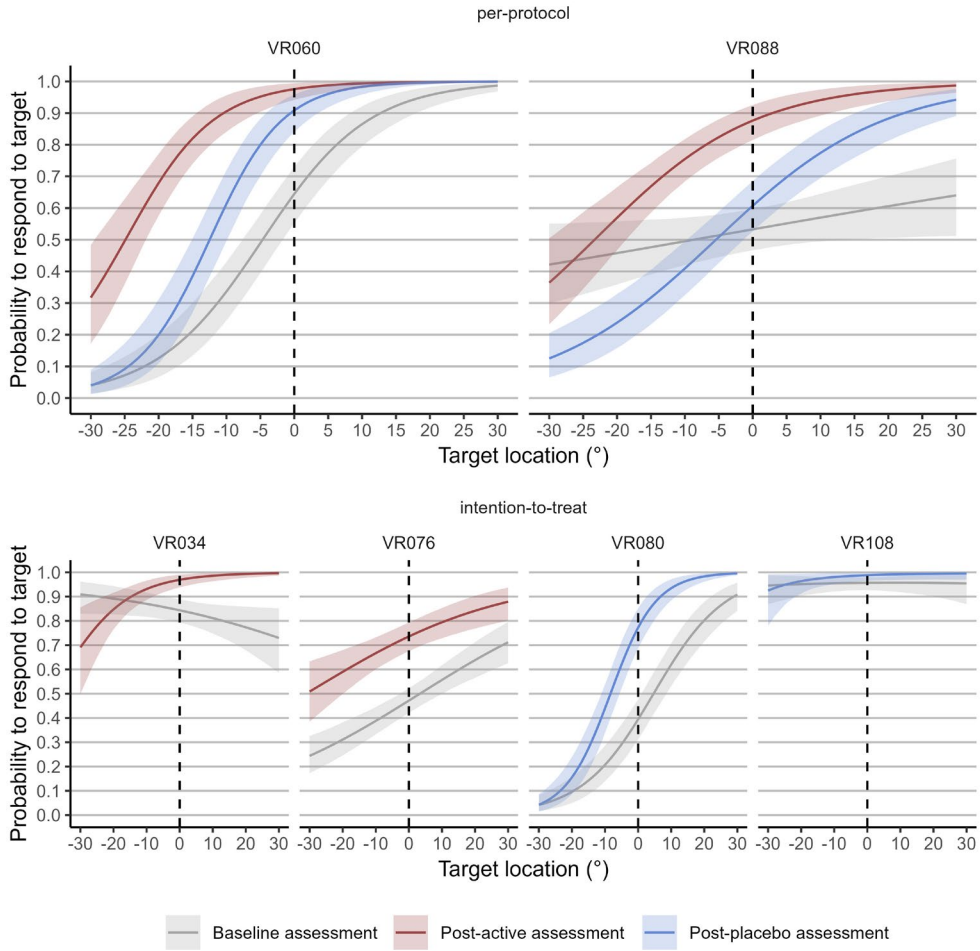

FIGURE 7 Estimated performance as a function of target location in the iVR environment for the baseline, post-placebo and post-active assessments for all cases. The ribbon represents the 95% credible interval of the posterior predictive distribution.

the Posner test for the left visual field during the active therapy phase relative to the placebo phase, in contrast to what was predicted (Figure 9). For right-sided targets, there was no statistically significant difference between the active and placebo therapy phases (mu: −.01, 95% CI = [−.02, .01]; sigma: .01, 95% CI [−.01, .03]) (Figure 9).

For VR034, there was no evidence in favour of faster response times over time for the left (mu: .01, 95% CI = [−.00, .02]; sigma: −.00, 95% CI = [−.01, .01]) and right visual field (mu: .00, 95% CI [−.01, .02]; sigma: .01, 95% CI [−.01, .02]).

For VR076, there was a significant reduction in mu of the lognormal distribution of response times for left-sided targets (−.08, 95% CI = [−.14, −.02], PP(H−) = 1), but no significant change in sigma of the lognormal distribution for left-sided targets (−.00, 95% CI = [−.04, .04]). There were no statistically significant changes in response times for right-sided targets (mu: −.01, 95% CI [−.04, .03]; sigma: −.02, 95% CI [−.04, .01]).

For VR080 there was no statistically significant change for the left visual field (mu: −.01, 95% CI = [−.04, .03]; sigma: −.00, 95% CI [−.03, .02]). The mu of the lognormal distribution of response times for the right visual field decreased (−.01, 95% CI [−.02, .00], PP(H−) = .95) and sigma increased (.02, 95% CI [.00, .04], PP(H+) = .99).

For VR108, there was no significant change in the response times for left-sided targets (mu: −.00, 95% CI [−.02, .01]; sigma: .01, 95% CI [−.01, .02]) (Figure 9). For the right visual field, there was an increase in mu of the lognormal distribution (.01, 95% CI [.00, .03], PP(H+) = .98) and a decrease in sigma (−.03, 95% CI [−.05, −.02], PP(H−) = 1).

Secondary outcome measures of spatial bias

In addition, four secondary outcome measures of spatial bias were used to assess the efficacy of the iVR training outside the iVR environment. The secondary outcome measures were not registered at intermediate therapy dosages but followed an interrupted time series design, often used in single-case research (Turner et al., 2020). For the per-protocol cases, changes in performance during the active therapy phase were compared to the preceding baseline-placebo phase. For intention-to-treat cases, performance was compared between a therapy phase and the two baseline measurements. We predicted a stronger improvement for left-sided targets as a result of the active therapy compared to the first phase (placebo phase for per-protocol patients and baseline phase for intention-to-treat patients). To integrate evidence across the secondary outcome measures, we used a Bayesian Evidence Synthesis method developed by Kuiper et al. (2013). This technique allows to synthesize evidence across multiple statistical tests.

For VR060, there was evidence in favour of a stronger improvement during the active therapy phase than the placebo phase on the computerized cancellation test (PP[6](H+) = 1) and on the Hearts cancellation test (PP(H+) = .61; PP(H0) = .31, PP(H−) = .08), but not on the Line Bisection test (PP(H+) = .02, PP(H0) = .12, PP(H−) = .86). Results for the CBS were inconclusive (PP(H+) = .31; PP(H0) = .42, PP(H−) = .28). Across tests, there was evidence for stronger improvements during the active phase than the placebo phase (PP(H+) = 1).

For VR088, there was evidence in favour of a reduced improvement during the active therapy phase than the placebo phase on the computerized cancellation test (PP(H+) = .01, PP(H0) = .06, PP(H−) = .94), the Hearts cancellation test (PP(H+) = .09, PP(H0) = .33, PP(H−) = .59) and on the Line Bisection test (PP(H+) = .07, PP(H0) = .29, PP(H−) = .64). Results for the CBS were inconclusive (PP(H+) = .28, PP(H0) = .42, PP(H−) = .31). Across tests, there was evidence for reduced improvements during the active therapy phase than the placebo phase (PP(H+) = .00, PP(H0) = .02, PP(H−) = .98).

For VR034, there was evidence in favour of more improvement during the active therapy than the baseline phase on the computerized cancellation test (PP(H+) = .69; PP(H0) = .26, PP(H−) = .06). However, this was not consistent with the Hearts cancellation test (PP(H+) = .06; PP(H0) = .28, PP(H−) = .66). The CBS was inconclusive (PP(H+) = .25; PP(H0) = .42, PP(H−) = .33). Across tests, evidence was inconclusive (PP(H+) = .21; PP(H0) = .56, PP(H−) = .23).

[6]PP = posterior probability.

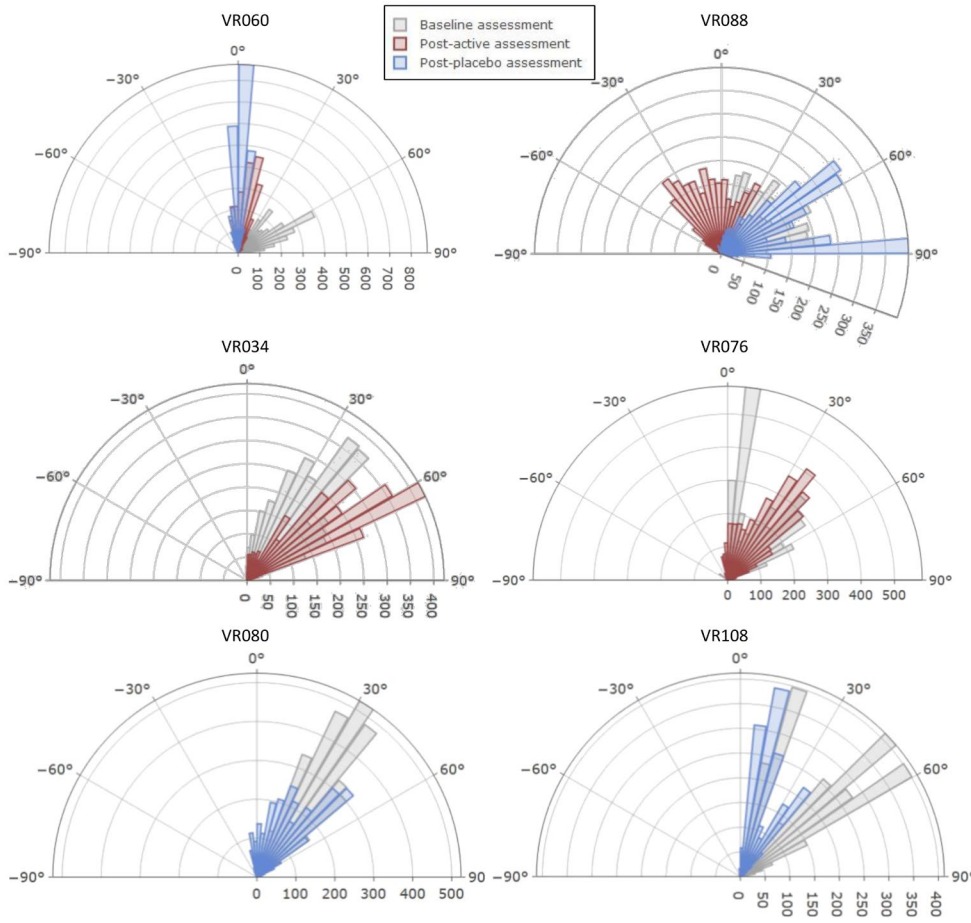

FIGURE 8 A wind rose diagram (frequencies) of head rotation (angles in degrees) during the IVR assessments for each patient (yaw rotation). 0° = straight ahead, <0° = left, >0° = right.

For VR076, there was evidence in favour of more improvements during the active therapy phase relative to the baseline phase on the computerized cancellation test (PP(H+) = .74, PP(H0) = .22, PP(H−) = .04). Results for the other tests did not provide clear evidence in favour of any of the three hypotheses (Hearts cancellation: PP(H+) = .48; PP(H0) = .38, PP(H−) = .13; Line Bisection: PP(H+) = .40; PP(H0) = .41, PP(H−) = .18). Across tests, there was evidence in favour of more improvements during the active therapy phase relative to the baseline phase (PP(H+) = .80, PP(H0) = .19, PP(H−) = .01).

For VR080, results for the Line Bisection test and CBS were inconclusive (Line Bisection: PP(H+) = .16, PP(H0) = .41, PP(H−) = .43; CBS: PP(H+) = .14, PP(H0) = .39, PP(H−) = .47). Across tests, evidence was inconclusive (PP(H+) = .06, PP(H0) = .41, PP(H−) = .53). For patient VR080 the cancellation tests were omitted from the analyses as the patient could not discriminate targets from distractors (see Appendix D). Patient VR108 could not be included in these analyses as the second baseline measurement was missing (see Appendix D).

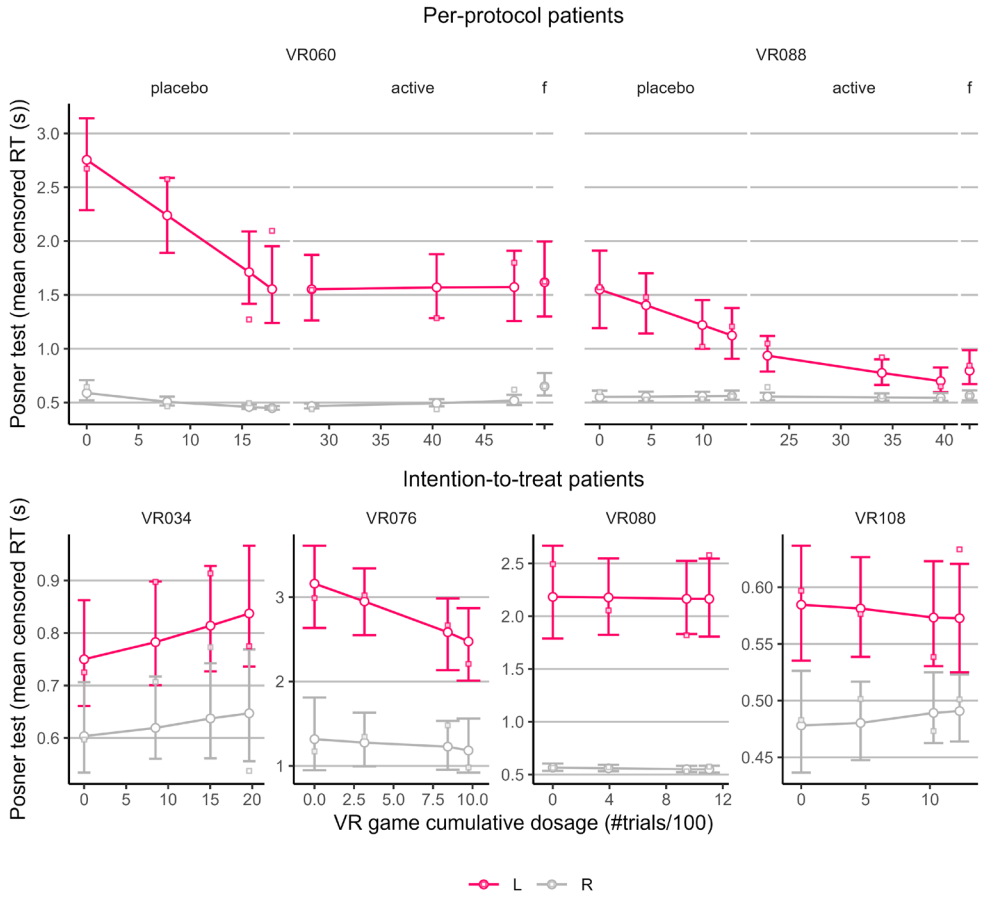

FIGURE 9 Performance on the primary outcome measure (i.e., Posner cueing test) of all patients as a function of cumulative treatment dosage in the VR game (number of trials divided by 100) and as a function of treatment condition (i.e., placebo versus active). For the two per-protocol-cases performance at the follow-up session (f) is also depicted. The round dots are the mean of the posterior predicted censored response times, while the square is the observed censored mean response times. Error bars are the 95% credible intervals of the posterior predictive distributions.

Secondary outcome measure of sustained attention

In addition to assessing the effect of the training on the spatial attention bias, we assessed its effect on non-spatial attention. To this end, we analysed performance on the SART (i.e., response times on go-trials and false alarms on no-go trials) over time (i.e. days since the start of observations). For the per-protocol cases, changes in performance during the active therapy phase were compared to the preceding baseline-placebo phase. For intention-to-treat cases, performance was compared between a therapy phase and the two baseline measurements.

For VR060, there was no statistically significant difference in the change of false alarms on the SART across time relative to the placebo phase (−.08, 95% CI [−.17, .00], PP(H−) = .97), nor for response times (Mean: 0, 95% CI [.00, .00], PP(H−) = .65; SD: .00, 95% CI [−.01, .01], PP(H−) = .20). For VR088, there was no statistically significant difference in the change of false alarms across time on the SART relative to the placebo phase (.02, 95% CI [−.05, .1], PP(H−) = .27), nor was there evidence in favour of an increased reduction over time in response times (Mean: .00, 95% CI [.00, .01], PP(H−) = .21; SD: −.01, 95% CI [−.02, .00], PP(H−) = .89).

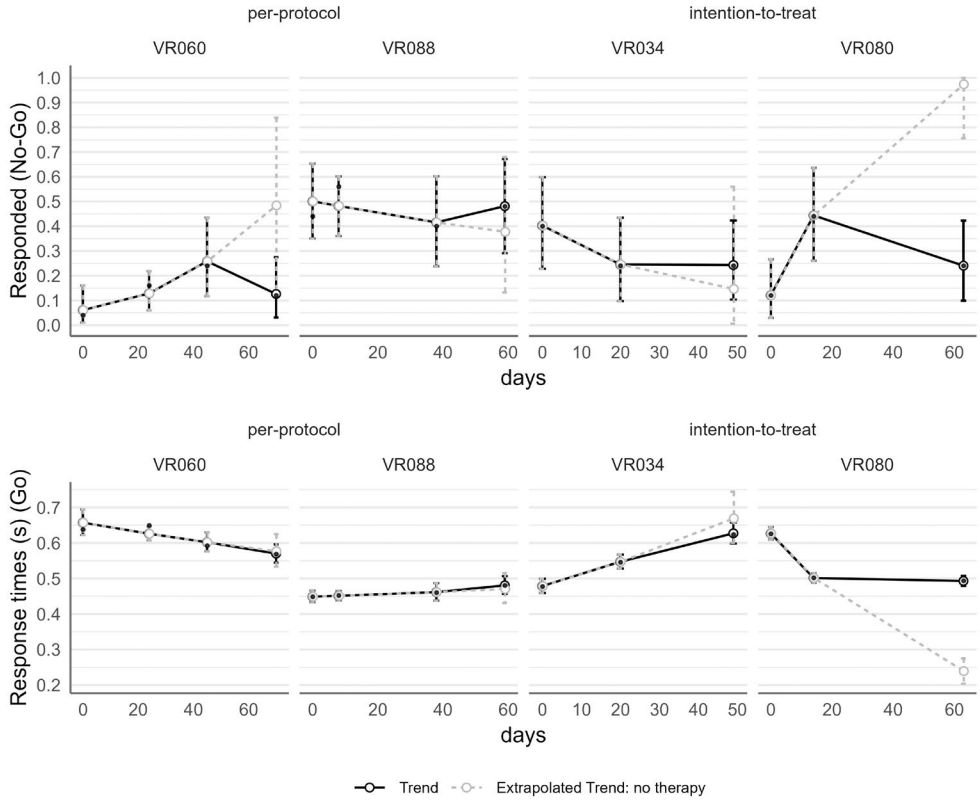

FIGURE 10 Results of the Sustained Attention to Response Task. Estimated probability of responses on no-go trials (top row) and estimated response times on go-trials (bottom row) across repeated assessments (days since start of observations). The grey dashed line is the extrapolated trend based on the first phase (before active therapy for per-protocol cases and before therapy for intention-to-treat cases). For cases VR076 and VR108, there were timepoints missing, and they were not further analysed. The open dot is the estimated mean value, the error bar is the 95% credible interval of the posterior predictions, and the full dot is the observed mean value.

For VR034, there was no statistically significant stronger improvement over time on the SART during the active therapy phase relative to the baseline phase (False alarms: PP(H−) = .22; Mean RTs: PP(H−) = .94; SD RTs: PP(H−) = .00). For VR080, there was evidence for a significant difference in change over time in false alarms during the placebo therapy phase relative to the baseline (−.16, 95% CI [−.28, −.05], PP(H−) = 1), but no evidence for increased improvement in response times (Mean: PP(H−) = .00; SD: PP(H−) = .41) (Figure 10). Patient VR108 could not be included in these analyses as the second baseline measurement was missing (see Appendix D).

Transfer: Effects inside the iVR environment versus outside the iVR environment

To evaluate the possibility that training effects were selective to the iVR environment but did not transfer to tests administered outside the iVR environment, we compared improvements from pre- (i.e., assessment before the post-active assessment) to post-active training between the iVR assessment and the computerized cancellation test. In case of a lack of transfer, we would expect stronger improvements for left targets on the iVR assessment compared to the computerized cancellation test. There was, however, no evidence in support of this hypothesis for patient VR034 (PP(H−) = .19), VR076 (PP(H−) = .14)

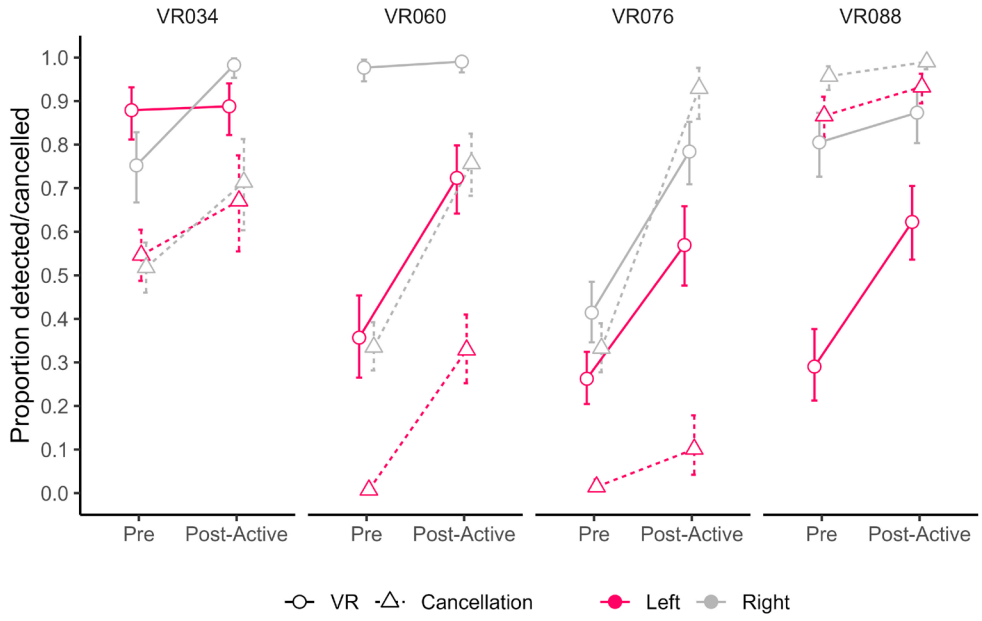

FIGURE 11 Estimated proportion detected (IVR) or cancelled (computerized cancellation test) targets as a function of time (pre- and post-active training), task (IVR or cancellation task) and visual field for each case completing active training. The error bars represent the 95% credible intervals of the posterior predictive distributions, and shapes represent the mean estimate.

and VR060 (PP(H−) = .00) (Figure 11). For patient VR088, evidence was in support of the hypothesis (−.63, 95% CI [−1.52, .27], PP(H−) = .92), but this may result from a high performance at the baseline assessment in the cancellation test (Figure 11). Detailed results of all regression models are reported in Appendix E; Tables E1–E6; Figures E1–E4.

DISCUSSION

Low feasibility of the planned and preregistered protocol

We initially planned to investigate the efficacy and feasibility of a novel iVR treatment for hemispatial neglect. Existing treatments focused on enhancing attentional orientation towards the left visual field using cueing (e.g., visual scanning therapy). However, these conventional approaches were not tailored to the patient's visuospatial attention distribution nor used head-contingent spatial cueing. We hypothesized that spatially biased head-contingent multisensory cueing would result in accelerated improvements in the visuospatial bias relative to unbiased head-contingent cueing. The study we planned (and preregistered) to test this hypothesis proved infeasible, which we now recognize as a key finding. While our goal was to include at least 8 stroke patients who met the per-protocol criteria, we were only able to recruit 2 per-protocol patients over a 2-year period. This shortfall was primarily due to the stringent preregistered in- and exclusion criteria (i.e., patients with left-sided neglect following a right-hemispheric stroke, as determined by a computerized cancellation test) and the demands of the intensive training and assessment protocol (i.e., 10 weeks, 5 sessions per week, 1-h sessions). Out of 110 screened patients, only 6 met the eligibility criteria, and the high drop-out rate (4 out of 6 patients) further compromised our ability to reach the study's predetermined endpoints. As a result, we made the decision to conclude the study prematurely and focus on discussing the low feasibility of our protocol as a central outcome.

One key challenge was that the prevalence of neglect in our screened sample was lower than anticipated, that is, 10% rather than the expected 30% in the subacute phase (<3 months post-stroke) (Esposito et al., 2020). This discrepancy may be partially due to our use of a more stringent diagnostic criterion, which classified patients as having neglect only if they showed a statistically significant difference in the proportion of cancelled targets between the left and the right visual field on a computerized cancellation task. This method offers higher specificity compared to many conventional diagnostic approaches, as it better distinguishes true spatial biases from non-spatial errors and reduces false positives often caused by multiple comparisons (Huygelier et al., 2021; Huygelier, Moore, et al., 2020). While this conservative criterion may have contributed to the lower prevalence, it is also possible that the decreasing severity of stroke played a role. Stroke severity and post-stroke disability have been declining over time, likely due to advancements in preventive healthcare and acute stroke care (Bernegger et al., 2022; Koton et al., 2018; Toyoda et al., 2022; Wafa et al., 2020). Consequently, the current prevalence of severe neglect may be lower than previously reported in the literature. These results underscore the importance of developing standardized and evidence-based criteria for diagnosing hemispatial neglect. Such criteria could include the utilization of standardized test batteries with minimal redundancy, the establishment of core outcome measures and the definition of sensitive and specific diagnostic cut-offs. By adopting these practices, we can achieve more consistent prevalence rates across various settings, improve the comparability of patient samples in research and ultimately pave the way for more effective clinical trials.

A second factor contributing to the low number of per-protocol patients was the planned frequency of sessions (47 sessions over a 10-week period), which proved unfeasible for stroke patients residing in a Belgian rehabilitation facility. The intense protocol was indeed difficult to accommodate alongside treatment as usual and required rest times. This highlights the need for more efficient training tools that can target multiple outcomes (e.g., motor functioning, visuospatial functioning) in a synergistic fashion. Such tools can be of interest to multiple therapeutic disciplines and could be more easily integrated into patients' rehabilitation schedule. Additionally, the current protocol was further intensified and lengthened by the high number of assessment sessions. Simplifying the protocol by reducing the number of outcome measures to an agreed-upon evidence-based test battery, focused on minimal redundancy, high sensitivity and specificity, would improve feasibility in future trials.

Importantly, the results showed that neglect patients are likely to adhere to an iVR-based neglect training protocol. All 6 patients completed at least 10 training sessions (in addition to multiple assessment sessions), demonstrating good adherence. Moreover, iVR training was generally well-tolerated by the patients. Although some patients reported mild physical discomfort, none experienced severe cybersickness or dropped-out due to discomfort. In fact, previous studies have shown that stroke patients tend to report fewer cybersickness complaints after iVR than before iVR exposure (Huygelier, Schraepen, et al., 2020). These findings suggest that VR training is a promising avenue for further exploration in the treatment of post-stroke neglect.

Improvements in performance and head orientation in the iVR environment

Although we were unable to perform the planned group analysis due to the study's low feasibility, we conducted single-case analyses to gain some preliminary insights into the potential rehabilitation effects.

Some of the results were in line with our expectations and suggest a potential positive effect of spatially biased iVR training. That is, we found that the two per-protocol patients detected more targets at locations further to the left of their head midline after the spatially biased (active) iVR training relative to the unbiased (placebo) iVR training inside the iVR environment. The iVR spatially biased training can thus reduce the head-contingent visuospatial bias inside the iVR environment. The latter requires further confirmation in larger samples. Interestingly, the results of our cases contrast with an earlier study that reported that neglect patients only improved on an endogenous Posner test (with central cue),

but not on an exogenous Posner test (peripheral cue) (Turgut et al., 2021). In our study, patients seemed to learn to orient attention towards peripheral visual targets in the left visual field after training with peripheral cues. This difference in results may be related to the use of multisensory rhythmic looming cues to promote attentional orientation towards the contralesional visual field. Previous studies have indeed shown that healthy individuals were better able to attend to visual signals when they were coupled to a rhythmic auditory signal (van Ee et al., 2009) and recent studies have shown beneficial effects of multisensory cues for rehabilitation of hemianopia and spatial neglect (Làdavas et al., 2020; Rowland et al., 2023).

In addition, we observed inconsistent effects of the iVR active training on head orientation. These inconsistencies are likely due to the absence of explicit instructions regarding head orientation and the head-contingent design, which allowed patients to complete the task regardless of their head orientation. Notably, in one case, the data suggested a positive and selective impact of the spatially biased iVR training on head orientation. This result suggests that the patient not only learnt to orient attention to the contralesional side of space when task-relevant but also showed a reduced preference to orient towards the ipsilesional side of space. In line with our results, a recent study demonstrated positive effects of active exploration training requiring patients to rotate their trunk towards the contralesional side on measures of spatial neglect (Stammler et al., 2023). The results of the current study and the study by Stammler et al. (2023) may suggest a bidirectional interaction between biases in spatial attention and biases in body orientation.

Noteworthy is the fact that effects on performance in the iVR environment were clear after 10 iVR sessions in which patients played the active iVR game for 5–6 h, completing more than 2500 repetitions of the iVR task. After this treatment dosage, there was still a considerable tendency to ignore left-sided stimuli in the iVR task. A higher therapy dosage (i.e., more task repetitions) may thus be required to increase therapeutic effects. Indeed, recent studies have confirmed that higher therapy dosages are effective in improving post-stroke outcomes such as arm functioning and visuospatial neglect (Chen et al., 2022; Lohse et al., 2014; Winstein et al., 2019). Therefore, an improved study protocol with a higher dosage of active therapy is recommended to adequately assess the efficacy of iVR spatially biased training.

No evidence for superior effects on measures outside the IVR environment

While the spatially biased iVR training appeared to have a selective effect *within* the iVR environment, we found no consistent evidence supporting a generalized treatment effect on non-iVR tasks. Moreover, due to the small per-protocol sample size, drawing firm conclusions on this matter remains challenging.

The lack of evidence for a superior effect may result from several factors. A first reason may be that the outcome measures outside the iVR environment only captured biases in a small area of the visual field, while the superior effect of the spatially biased iVR training was mostly evident for more eccentric target locations in the iVR task. Thus, a potential superior effect of the iVR spatially biased training relative to the unbiased training may not have been detectable within this small area of the visual field measured by the non-iVR tasks.

The challenge of achieving task-specific training effects instead of broader, generalized effects has been a persistent issue. Indeed, research on executive function training has, for instance, demonstrated near-transfer effects (impact on closely related tasks) but no far-transfer effects (impact on tasks that are loosely related to the trained task) (Smid et al., 2020). Although all outcome measures in the current study aim to measure visuospatial attention biases, the cognitive processes recruited by the tasks may vary too extensively to result in generalized training effects. Although training of higher cognitive functions has often failed to provide convincing evidence for generalizable effects (Smid et al., 2020), studies on cats have provided clear evidence for the possibility to restore visual functions using multisensory training after inducing lesions (Stein & Rowland, 2020). These studies have demonstrated that repetitive functional training can result in restorative treatment effects and suggest that rehabilitation

should not solely focus on compensatory strategies for ameliorating visuo-perceptual deficits. Studies have also identified positive effects of audiovisual stimulation for hemianopia and neglect after stroke (Tinga et al., 2015), warranting further research into multisensory restorative rehabilitation for hemianopia and neglect.

The lack of evidence supporting a superior effect of active iVR training on non-iVR outcome measures should not be interpreted as evidence that the iVR training is inherently ineffective. In our comparison, we focused on two training conditions differing solely in the spatial distribution of cueing. This comparison does not shed light on potential holistic effects of iVR training in contrast to conventional therapy. The latter is plausible as previous studies demonstrated that non-spatial attention training had positive effects on spatial biases (Van Vleet & DeGutis, 2013; Vleet et al., 2020). Thus, the spatially unbiased (placebo) iVR training may have had a positive impact on the patient's neglect symptoms. It is thus important to further evaluate the efficacy of the iVR training as a whole relative to a dosage-matched inactive condition.

Large interindividual variability

The current study also clarified that selecting patients for the iVR training based on the performance on the computerized cancellation test was not ideal. That is, two patients (i.e., VR034 and VR108) did not demonstrate clear spatial biases in task performance on the iVR task at baseline, although they did have a statistically significant bias on the computerized cancellation test. These two cases were unlikely to benefit from the spatially biased iVR training given their high baseline performance.

There may be several reasons why spatial biases on the computerized cancellation tasks did not correspond with spatial biases on the iVR task for all patients. A first reason is the fact that the cancellation task involves a stronger motor component than the iVR task. That is, to mark a target on the computerized cancellation test, patients have to move their ipsilesional limb towards contralesional space (although they do not need to cross the midline as in a pen and paper cancellation test). Cancellation tests likely represent a mix of biases in spatial attention and directional hypokinesia (Sapir et al., 2007). This motor component is absent in the iVR task. Consistent with this interpretation, patient VR108 had a lesion affecting the putamen, which has been linked to directional hypokinesia in previous research (Sapir et al., 2007). A second reason for differences between tasks may be the differential involvement of top-down versus bottom-up attention in the cancellation test versus the iVR task. That is, in the cancellation test, targets are static, and patients have to guide their visual search towards the left visual field. In the iVR task, targets are brief onsets that can capture attention in a stimulus-driven fashion. Thus, it may be possible that these two patients had a selective deficit in top-down attention and not in stimulus-driven attention. However, the latter interpretation is not consistent with studies reporting more severe deficits in exogenous than endogenous attention in neglect (Bartolomeo & Chokron, 2002; Losier & Klein, 2001) and with the fact that the lesion of VR034 did not selectively affect the dorsal attention network, which is considered to be important for top-down attention (Corbetta & Shulman, 2011). Another possibility may be related to the fact that targets were presented within peri-personal space (i.e., within reaching space) in the cancellation test and extra-personal space (i.e., outside of reaching space) in the iVR task. Studies have indeed reported that certain patients have selective deficits in peri- or extra-personal space (Aimola et al., 2012; Cowey et al., 1994; Stoep et al., 2013; Ten Brink et al., 2019). Last, the differences between tasks may merely be related to the sensitivity of the tests rather than the specific cognitive processes involved in the tests. That is, the time pressure in the iVR task (3 s response window) may not have been sufficient to detect a spatial bias. Indeed, in the Posner test, these two patients were much faster to detect targets in the left visual field (responding in less than 1 s) in comparison to the other patients (responding in 2–3 s). These results highlight the need for fundamental research to clarify why patient performance on tests of visuospatial attention is often inconsistent. Such

research is crucial for improving evidence-based guidelines for neglect assessment and conducting clinical studies that can offer clear, reliable evidence regarding treatment effects.

Going from group-based randomized controlled trials to single-case studies?

Despite several limitations, the current study also has several strengths. First, our study protocol and analysis plan were pre-registered, preventing us from deviating from our initial study plan. Although the latter offered us little flexibility to adjust our study protocol, it does increase insight into which clinical trial designs for neglect are feasible. This offers maximal transparency and may guide others in designing rehabilitation studies in the future.

Second, we used a longitudinal design paired with computerized tests to quantify neglect recovery over time. The latter enabled us to obtain precise estimates of spatial bias at the single-case level and consider patients' individual recovery trajectories. Our single-case data clearly demonstrates pronounced interindividual differences in trends over time and severity of spatial biases between patients. These results suggest that group-level analysis on small samples likely masks important interindividual differences and may be sensitive to outliers. Deciding which treatment to implement for neglect merely based on statistically significant group-level effects and generalizing the grand average to all patients is unlikely to maximize the treatment effect, as Lee Cronbach already noted in 1957 (Cronbach, 1957). Single-case interrupted time series designs, frequently used when randomized controlled trials are not feasible (Turner et al., 2020), may thus be an important way forward for neglect rehabilitation research. Rehabilitation needs to be effective at the single-case level, and understanding who can benefit from which treatment at what point in time is the way forward. To this end, future neurorehabilitation studies should embrace rather than hide interindividual heterogeneity and embark on fundamental studies that can identify why these differences occur.

Conclusion

In conclusion, our results showed that a longitudinal, cross-over, placebo-controlled study to investigate iVR rehabilitation was not feasible. Low enrolment and high drop-out rates prevented us from conclusively evaluating the efficacy (and potential superiority) of spatially biased iVR training compared to spatially unbiased iVR training. However, the preliminary single-case analyses revealed several interesting findings that warrant further research. First, the spatially biased head-contingent multisensory cueing in our 2 per-protocol cases resulted in notable improvements in task performance within the iVR environment. These effects in the iVR environment were evident after 10 sessions. Our study, however, did not provide consistent evidence supporting superior treatment effects of the spatially biased iVR training on measures outside the iVR environment. This may be attributed to the specificity of outcome measures or a lack of transfer of training effects and requires more investigation.

Although the study encountered challenges with feasibility, it is important to emphasize that the iVR training demonstrated safety. Based on our findings, future research should investigate the broader effects of iVR training compared to conventional therapy, using an optimized study protocol. More generally, the development of evidence-based, standardized guidelines for neglect assessment is essential to advance the field of neglect rehabilitation.

AUTHOR CONTRIBUTIONS

Hanne Huygelier: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; validation; visualization; writing – original draft; writing – review and editing. **Nora Tuts:** Data curation; project administration; writing – review and editing. **Karla Michiels:** Data curation; writing – review and editing. **Eline Note:** Data curation; writing – review and editing. **Fabienne Schillebeeckx:** Data curation; writing – review and editing.

Jos Tournoy: Data curation; writing – review and editing. **Vero Vanden Abeele:** Conceptualization; funding acquisition; supervision; writing – review and editing. **Raymond van Ee:** Conceptualization; funding acquisition; supervision; writing – review and editing. **Céline R. Gillebert:** Conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

This work was funded by a research grant of the Flemish Fund for Scientific Research (FWO) awarded to H.H. (1249923N), R.v.E. and V.V.A. (G078915N) and C.R.G. (G0H7718N, G0B6622N). Equipment used in this study were funded by a Hercules infrastructure grant (AKUL/19/05). We would like to acknowledge the contributions of Ayten Köksal, Jelle Loyaerts and Elise Palmans in data collection for this study. We would like to acknowledge that we used chat-GPT version 3.5 as a writing assistance for the discussion section. OpenAI. (2023). ChatGPT (version 3.5) [Large language model]. https://chat.openai.com/chat.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to disclose.

DATA AVAILABILITY STATEMENT

The data, analysis scripts and checks on model fit are publicly available on FigShare (https://doi.org/10.6084/m9.figshare.24873813).

ORCID

Hanne Huygelier https://orcid.org/0000-0002-5177-1193
Vero Vanden Abeele https://orcid.org/0000-0002-3031-9579
Raymond van Ee https://orcid.org/0000-0002-1134-8665
Céline R. Gillebert https://orcid.org/0000-0001-6686-7262

REFERENCES

Aimola, L., Schindler, I., Simone, A. M., & Venneri, A. (2012). Near and far space neglect: Task sensitivity and anatomical substrates. *Neuropsychologia*, *50*(6), 1115–1123. https://doi.org/10.1016/j.neuropsychologia.2012.01.022

Andersen, S. W., & Millen, B. A. (2013). On the practical application of mixed effects models for repeated measures to clinical trial data. *Pharmaceutical Statistics*, *12*(1), 7–16. https://doi.org/10.1002/pst.1548

Appel, L., Appel, E., Bogler, O., Wiseman, M., Cohen, L., Ein, N., Abrams, H. B., & Campos, J. L. (2020). Older adults with cognitive and/or physical impairments can benefit from immersive virtual reality experiences: A feasibility study. *Frontiers in Medicine*, *6*, 329. https://doi.org/10.3389/fmed.2019.00329

Arns, L. L., & Cerney, M. M. (2005). The relationship between age and incidence of cybersickness among immersive environment users. *IEEE Proceedings. VR 2005. Virtual Reality*, *2005*, 267–268. https://doi.org/10.1109/VR.2005.1492788

Azouvi, P., Olivier, S., de Montety, G., Samuel, C., Louis-Dreyfus, A., & Tesio, L. (2003). Behavioral assessment of unilateral neglect: Study of the psychometric properties of the Catherine Bergego scale. *Archives of Physical Medicine and Rehabilitation*, *84*(1), 51–57. https://doi.org/10.1053/apmr.2003.50062

Bailey, M. J., Riddoch, M. J., & Crome, P. (2004). Test–retest stability of three tests for unilateral visual neglect in patients with stroke: Star cancellation, line bisection, and the baking tray task. *Neuropsychological Rehabilitation*, *14*(4), 403–419. https://doi.org/10.1080/09602010343000282

Bartolomeo, P., & Chokron, S. (2002). Orienting of attention in left unilateral neglect. *Neuroscience & Biobehavioral Reviews*, *26*(2), 217–234. https://doi.org/10.1016/S0149-7634(01)00065-3

Bernegger, A., Mikšová, D., Posekany, A., Krebs, S., Ferrari, J., Greisenegger, S., Gattringer, T., Lang, W., Kiechl, S., Sykora, M., Mutzenbach, J. S., Bubel, N., Millesi, K., Katzenschlager, R., Torma, S., Krstic, M., Gruber, F., Vosko, M. R., Brunner, C., … The Austrian Stroke Unit Registry Collaborators. (2022). Time trends in stroke severity in the years 2005 to 2020: Results from the Austrian Stroke Unit Registry. *Journal of Neurology*, *269*(8), 4396–4403. https://doi.org/10.1007/s00415-022-11079-x

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the hospital anxiety and depression scale: An updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69–77. https://doi.org/10.1016/S0022-3999(01)00296-3

Bowen, A., Hazelton, C., Pollock, A., & Lincoln, N. B. (2013). Cognitive rehabilitation for spatial neglect following stroke. In The Cochrane Collaboration (Ed.), *Cochrane database of systematic reviews*. John Wiley & Sons, Ltd. https://doi.org/10.1002/14651858.CD003586.pub3

Burke, J. W., McNeill, M. D. J., Charles, D. K., Morrow, P. J., Crosbie, J. H., & McDonough, S. M. (2009). Optimising engagement for stroke rehabilitation using serious games. *The Visual Computer*, *25*(12), 1085–1099. https://doi.org/10.1007/s00371-009-0387-4

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Buxbaum, L. J., Ferraro, M. K., Veramonti, T., Farne, A., Whyte, J., Ladavas, E., Frassinetti, F., & Coslett, H. B. (2004). Hemispatial neglect subtypes, neuroanatomy, and disability. *Neurology*, *62*(5), 749–756. https://doi.org/10.1212/01.WNL.0000113730.73031.F4

Cazzoli, D., Müri, R. M., Schumacher, R., von Arx, S., Chaves, S., Gutbrod, K., Bohlhalter, S., Bauer, D., Vanbellingen, T., Bertschi, M., Kipfer, S., Rosenthal, C. R., Kennard, C., Bassetti, C. L., & Nyffeler, T. (2012). Theta burst stimulation reduces disability during the activities of daily living in spatial neglect. *Brain: A Journal of Neurology*, *135*, 3426–3439. https://doi.org/10.1093/brain/aws182

Checketts, M., Mancuso, M., Fordell, H., Chen, P., Hreha, K., Eskes, G. A., Vuilleumier, P., Vail, A., & Bowen, A. (2020). Current clinical practice in the screening and diagnosis of spatial neglect post-stroke: Findings from a multidisciplinary international survey. *Neuropsychological Rehabilitation*, *31*(9), 1–32. https://doi.org/10.1080/09602011.2020.1782946

Chen, P., Chen, C. C., Hreha, K., Goedert, K. M., & Barrett, A. M. (2015). Kessler Foundation Neglect Assessment Process uniquely measures spatial neglect during activities of daily living. *Archives of Physical Medicine and Rehabilitation*, *96*(5), 869–876.e1. https://doi.org/10.1016/j.apmr.2014.10.023

Chen, P., Hreha, K., Gonzalez-Snyder, C., Rich, T. J., Gillen, R. W., Parrott, D., & Barrett, A. M. (2022). Impacts of prism adaptation treatment on spatial neglect and rehabilitation outcome: Dosage matters. *Neurorehabilitation and Neural Repair*, *36*(8), 500–513. https://doi.org/10.1177/15459683221107891

Chen, P., Pitteri, M., Gillen, G., & Ayyala, H. (2018). Ask the experts how to treat individuals with spatial neglect: A survey study. *Disability and Rehabilitation*, *40*(22), 2677–2691. https://doi.org/10.1080/09638288.2017.1347720

Contrino, B., & Lazic, S. E. (2017). *BayesCombo: Bayesian evidence combination* (1.0) [Computer software]. https://cran.r-project.org/web/packages/BayesCombo/index.html

Corbetta, M., & Shulman, G. L. (2011). Spatial neglect and attention networks. *Annual Review of Neuroscience*, *34*, 569–599. https://doi.org/10.1146/annurev-neuro-061010-113731

Cowey, A., Small, M., & Ellis, S. (1994). Left visuo-spatial neglect can be worse in far than in near space. *Neuropsychologia*, *32*(9), 1059–1066.

Cremers, J., & Klugkist, I. (2018). One direction? A tutorial for circular data analysis using R with examples in cognitive psychology. *Frontiers in Psychology*, *9*, 2040. https://doi.org/10.3389/fpsyg.2018.02040

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943

Davis, S., Nesbitt, K., & Nalivaiko, E. (2015). Comparing the onset of cybersickness using the oculus rift and two virtual roller coasters. *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015)*, *167*, 12.

de Haan, B., Clas, P., Juenger, H., Wilke, M., & Karnath, H.-O. (2015). Fast semi-automated lesion demarcation in stroke. *NeuroImage: Clinical*, *9*, 69–74. https://doi.org/10.1016/j.nicl.2015.06.013

Demeyere, N., & Gillebert, C. (2017). Dissociating ego- and allocentric neglect after stroke: Prevalence, laterality and outcome predictors. *bioRxiv*, 200386. https://doi.org/10.1101/200386

Demeyere, N., Gillebert, C., Loftus, L., & Humphreys, G. (2015). Egocentric and allocentric neglect after right and left hemisphere lesions in a large scale neglect study of acute stroke patients: Prevalence and recovery. *Journal of Vision*, *15*(12), 179. https://doi.org/10.1167/15.12.179

Demeyere, N., & Gillebert, C. R. (2019). Ego- and allocentric visuospatial neglect: Dissociations, prevalence, and laterality in acute stroke. *Neuropsychology*, *33*(4), 490–498. https://doi.org/10.1037/neu0000527

Demeyere, N., Riddoch, M. J., Slavkova, E. D., Bickerton, W., & Humphreys, G. W. (2015). The Oxford Cognitive Screen (OCS): Validation of a stroke-specific short cognitive screening tool. *Psychological Assessment*, *27*(3), 883–894. https://doi.org/10.1037/pas0000082

Dermody, G., Whitehead, L., Wilson, G., & Glass, C. (2020). The role of virtual reality in improving health outcomes for community-dwelling older adults: Systematic review. *Journal of Medical Internet Research*, *22*(6), e17331. https://doi.org/10.2196/17331

Esposito, E., Shekhtman, G., & Chen, P. (2020). Prevalence of spatial neglect post-stroke: A systematic review. *Annals of Physical and Rehabilitation Medicine*, *64*, 101459. https://doi.org/10.1016/j.rehab.2020.10.010

Evald, L., Wilms, I., & Nordfang, M. (2020). Assessment of spatial neglect in clinical practice: A nationwide survey. *Neuropsychological Rehabilitation*, *31*(9), 1–16. https://doi.org/10.1080/09602011.2020.1778490

Foerster, R. M., Poth, C. H., Behler, C., Botsch, M., & Schneider, W. X. (2016). Using the virtual reality device oculus rift for neuropsychological assessment of visual processing capabilities. *Scientific Reports*, *6*, 37016. https://doi.org/10.1038/srep37016

Fordell, H., Bodin, K., Bucht, G., & Malm, J. (2011). A virtual reality test battery for assessment and screening of spatial neglect. *Acta Neurologica Scandinavica*, *123*(3), 167–174. https://doi.org/10.1111/j.1600-0404.2010.01390.x

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*(4), 733–760.

Grattan, E. S., Skidmore, E. R., & Woodbury, M. L. (2018). Examining anosognosia of neglect. *OTJR: Occupation, Participation and Health*, *38*(2), 113–120. https://doi.org/10.1177/1539449217747586

Griffis, J. C., Metcalf, N. V., Corbetta, M., & Shulman, G. L. (2021). Lesion quantification toolkit: A MATLAB software tool for estimating grey matter damage and white matter disconnections in patients with focal brain lesions. *NeuroImage: Clinical*, *30*, 102639. https://doi.org/10.1016/j.nicl.2021.102639

Guo, J., Xu, X., Ying, Z., & Zhang, S. (2022). Modeling not-reached items in timed tests: A response time censoring approach. *Psychometrika*, *87*(3), 835–867. https://doi.org/10.1007/s11336-021-09810-0

Halligan, P. W., Cockburn, J., & Wilson, B. A. (1991). The behavioural assessment of visual neglect. *Neuropsychological Rehabilitation*, *1*(1), 5–32. https://doi.org/10.1080/09602019108401377

Harvey, M. (2019). *Neglect rehabilitation: Strength and challenges of treatments using non-invasive brain stimulation*. 7th scientific meeting of the Federation of the European Societies of Neuropsychology, Milan, Italy. https://www.fesn2019.it/relations/Harvey_Monika_FESN_2019.pdf

Husain, M., & Rorden, C. (2003). Non-spatially lateralized mechanisms in hemispatial neglect. *Nature Reviews Neuroscience*, *4*(1), 26–36. https://doi.org/10.1038/nrn1005

Huygelier, H., Gillebert, C. R., & Moors, P. (2021). The value of Bayesian methods for accurate and efficient neuropsychological assessment. *Journal of the International Neuropsychological Society*, *28*, 1–12. https://doi.org/10.1017/S1355617721001120

Huygelier, H., Gillebert, C. R., van Ee, R., & Vanden Abeele, V. (2017). The design of a virtual reality game for stroke-induced attention deficits. In *Extended abstracts publication of the annual symposium on computer-human interaction in play*, 223–230. ACM/IEEE. https://doi.org/10.1145/3130859.3131308

Huygelier, H., Moore, M. J., Demeyere, N., & Gillebert, C. R. (2020). Non-spatial impairments affect false positive neglect diagnosis based on cancellation tasks. *Journal of the International Neuropsychological Society*, *26*, 668–678. https://doi.org/10.1017/S1355617720000041

Huygelier, H., Schraepen, B., Demeyere, N., & Gillebert, C. R. (2019). The Dutch version of the Oxford Cognitive Screen (OCS-NL): Normative data and their association with age and socio-economic status. *Aging, Neuropsychology, and Cognition*, *27*, 1–22. https://doi.org/10.1080/13825585.2019.1680598

Huygelier, H., Schraepen, B., Lafosse, C., Vaes, N., Schillebeeckx, F., Michiels, K., Note, E., Abeele, V. V., van Ee, R., & Gillebert, C. R. (2020). An immersive virtual reality game to train spatial attention orientation after stroke: A feasibility study. *Applied Neuropsychology: Adult*, *29*(5), 1–21. https://doi.org/10.1080/23279095.2020.1821030

Huygelier, H., Schraepen, B., van Ee, R., Vanden Abeele, V., & Gillebert, C. R. (2019). Acceptance of immersive head-mounted virtual reality in older adults. *Scientific Reports*, *9*(1), 4519. https://doi.org/10.1038/s41598-019-41200-6

Jammalamadaka, S. R., Guerrier, S., & Mangalam, V. (2021). A two-sample nonparametric test for circular data- its exact distribution and performance. *Sankhya. Series B (2008)*, *83*(Suppl 1), 140–166. https://doi.org/10.1007/s13571-020-00244-9

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the Americal Statistical Association*, *90*(430), 773–795.

Kerkhoff, G., & Schenk, T. (2012). Rehabilitation of neglect: An update. *Neuropsychologia*, *50*(6), 1072–1079. https://doi.org/10.1016/j.neuropsychologia.2012.01.024

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M., & Horwitz, R. I. (1999). Stratified randomization for clinical trials. *Journal of Clinical Epidemiology*, *52*(1), 19–26. https://doi.org/10.1016/S0895-4356(98)00138-3

Kolodziej, W., & Gillebert, C. R. (2018). *Diagnostiek en revalidatie van hemispatiaal neglect in Vlaanderen en Nederland*. KU Leuven.

Koton, S., Geva, D., Streifler, J. Y., Harnof, S., Pougach, Y., Azrilin, O., Hadar, S., Bornstein, N. M., & Tanne, D. (2018). Declining rate and severity of hospitalized stroke from 2004 to 2013. *Stroke*, *49*(6), 1348–1354. https://doi.org/10.1161/STROKEAHA.117.019822

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Technological competence is a pre-condition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. *Frontiers in Human Neuroscience*, *13*, 1–17. https://doi.org/10.3389/fnhum.2019.00342

Kuiper, R. M., Buskens, V., Raub, W., & Hoijtink, H. (2013). Combining statistical evidence from several studies: A method using Bayesian updating and an example from research on trust problems in social and economic exchange. *Sociological Methods & Research*, *42*(1), 60–81. https://doi.org/10.1177/0049124112464867

Làdavas, E., Tosatto, L., & Bertini, C. (2020). Behavioural and functional changes in neglect after multisensory stimulation. *Neuropsychological Rehabilitation*, *32*(5), 1–28. https://doi.org/10.1080/09602011.2020.1786411

Laplane, D., & Degos, J. D. (1983). Motor neglect. *Journal of Neurology, Neurosurgery, and Psychiatry*, *46*(2), 152–158.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence Teleoperators and Virtual Environments*, *10*(3), 282–297. https://doi.org/10.1162/105474601300343612

Lohse, K. R., Lang, C. E., & Boyd, L. A. (2014). Is more better? Using metadata to explore dose–response relationships in stroke rehabilitation. *Stroke*, *45*(7), 2053–2058. https://doi.org/10.1161/STROKEAHA.114.004695

Losier, B. J. W., & Klein, R. M. (2001). A review of the evidence for a disengage deficit following parietal lobe damage. *Neuroscience & Biobehavioral Reviews*, *25*(1), 1–13. https://doi.org/10.1016/S0149-7634(00)00046-4

Luauté, J., Halligan, P., Rode, G., Rossetti, Y., & Boisson, D. (2006). Visuo-spatial neglect: A systematic review of current interventions and their effectiveness. *Neuroscience & Biobehavioral Reviews*, *30*(7), 961–982. https://doi.org/10.1016/j.neubiorev.2006.03.001

Lund, U., Agostinelli, C., Arai, H., Gagliardi, A., Garcia-Portugues, E., Giunchi, D., Irisson, J.-O., Pocernich, M., & Rotolo, F. (2023). *Circular* (0.5-0) [R].

Luvizutto, G. J., Bazan, R., Braga, G. P., de Lima Resende, L. A., Bazan, S. G. Z., & Dib, R. E. (2015). Pharmacological interventions for unilateral spatial neglect after stroke. *Cochrane Database of Systematic Reviews*, *11*, CD010882. https://doi.org/10.1002/14651858.CD010882.pub2

Machner, B., Mah, Y.-H., Gorgoraptis, N., & Husain, M. (2012). How reliable is repeated testing for hemispatial neglect? Implications for clinical follow-up and treatment trials. *Journal of Neurology, Neurosurgery, and Psychiatry*, *83*(10), 1032–1034. https://doi.org/10.1136/jnnp-2012-303296

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58. https://doi.org/10.1080/02701367.1989.10607413

McIntosh, R. D. (2017). *The end of the line: Antagonistic attentional weightings in unilateral neglect* [Article in Press]. Scopus. https://doi.org/10.1016/j.cortex.2017.07.011

McIntosh, R. D., Ietswaart, M., & Milner, A. D. (2017). Weight and see: Line bisection in neglect reliably measures the allocation of attention, but not the perception of length. *Neuropsychologia*, *106*, 146–158. https://doi.org/10.1016/j.neuropsychologia.2017.09.014

McIntosh, R. D., Schindler, I., Birchall, D., & Milner, A. D. (2005). Weights and measures: A new look at bisection behaviour in neglect. *Cognitive Brain Research*, *25*(3), 833–850. https://doi.org/10.1016/j.cogbrainres.2005.09.008

Milner, A. D., & Harvey, M. (1994). Toward a taxonomy of spatial neglect. *Neuropsychological Rehabilitation*, *4*(2), 177–181.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*, c869. https://doi.org/10.1136/bmj.c869

Nadarajah, M., Mazlan, M., & Goh, H. (2017). Test-retest reliability, internal consistency and concurrent validity of fatigue severity scale in measuring post-stroke fatigue. *European Journal of Physical and Rehabilitation Medicine*, *53*(5), 7.

National Research Council (US) Panel on Handling Missing Data in Clinical Trials. (2010). *The prevention and treatment of missing data in clinical trials*. National Academies Press (US). http://www.ncbi.nlm.nih.gov/books/NBK209904/

Negrini, S., Arienti, C., & Kiekens, C. (2020). Usual care: The big but unmanaged problem of rehabilitation evidence. *The Lancet*, *395*(10221), 337. https://doi.org/10.1016/S0140-6736(19)32553-X

Nijboer, T. C., Kollen, B. J., & Kwakkel, G. (2013). Time course of visuospatial neglect early after stroke: A longitudinal cohort study. *Cortex*, *49*(8), 2021–2027.

Oculus Rift | Oculus. (n.d.). https://www.oculus.com/rift/#oui-csl-rift-games=mages-tale

Peirce, J. W. (2007). PsychoPy—Psychophysics software in python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Plechatá, A., Sahula, V., Fayette, D., & Fajnerová, I. (2019). Age-related differences with immersive and non-immersive virtual reality in memory assessment. *Frontiers in Psychology*, *10*, 1330. https://doi.org/10.3389/fpsyg.2019.01330

Porcino, T. M., Clua, E., Trevisan, D., Vasconcelos, C. N., & Valente, L. (2017). Minimizing cyber sickness in head mounted display systems: Design guidelines and applications. *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*, 1–6. https://doi.org/10.1109/SeGAH.2017.7939283

Rizzo, A. A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation*, *14*(1–2), 207–239.

Robertson, I. H. (2001). Do we need the "lateral" in unilateral neglect? Spatially nonselective attention deficits in unilateral neglect and their implications for rehabilitation. *NeuroImage*, *14*(1), S85–S90.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758. https://doi.org/10.1016/S0028-3932(97)00015-8

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization. *NeuroImage*, *61*(4), 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rowland, B. A., Bushnell, C. D., Duncan, P. W., & Stein, B. E. (2023). Ameliorating hemianopia with multisensory training. *Journal of Neuroscience*, *43*(6), 1018–1026. https://doi.org/10.1523/JNEUROSCI.0962-22.2022

Salazar, A. P. S., Vaz, P. G., Marchese, R. R., Stein, C., Pinto, C., & Pagnussat, A. S. (2018). Noninvasive brain stimulation improves hemispatial neglect after stroke: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation*, *99*(2), 355–366.e1. https://doi.org/10.1016/j.apmr.2017.07.009

Sapir, A., Kaplan, J. B., He, B. J., & Corbetta, M. (2007). Anatomical correlates of directional hypokinesia in patients with hemispatial neglect. *The Journal of Neuroscience*, *27*(15), 4045–4051. https://doi.org/10.1523/JNEUROSCI.0041-07.2007

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*, 322–339. https://doi.org/10.1037/met0000061

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a science of effective cognitive training. *Current Directions in Psychological Science*, *29*(6), 531–537. https://doi.org/10.1177/0963721420951599

Stammler, B., Flammer, K., Schuster, T., Lambert, M., Neumann, O., Lux, M., Matuz, T., & Karnath, H.-O. (2023). *Spatial neglect therapy with the augmented reality app "Negami" for active exploration training; A randomized controlled trial* (p. 2023.03.26.23287770). *medRxiv*. https://doi.org/10.1101/2023.03.26.23287770

Stanney, K. M., & Hash, P. (1998). Locus of user-initiated control in virtual environments: Influences on cybersickness. *Presence Teleoperators and Virtual Environments*, *7*(5), 447–459. https://doi.org/10.1162/105474698565848

Stein, B. E., & Rowland, B. A. (2020). Using superior colliculus principles of multisensory integration to reverse hemianopia. *Neuropsychologia*, *141*, 107413. https://doi.org/10.1016/j.neuropsychologia.2020.107413

Stoep, N. V., Visser-Meily, J. M. A., Kappelle, L. J., Kort, P. L. M., Huisman, K. D., Eijsackers, A. L. H., Kouwenhoven, M., Stigchel, S. V., & Nijboer, T. C. W. (2013). Exploring near and far regions of space: Distance-specific visuospatial neglect after stroke. *Journal of Clinical and Experimental Neuropsychology*, *35*(8), 799–811. https://doi.org/10.1080/13803395.2013.824555

Sturm, W., Thimm, M., Binkofski, F., Horoufchin, H., Fink, G. R., Küst, J., Karbe, H., & Willmes, K. (2013). Combined space and alertness related therapy of visual hemineglect: Effect of therapy frequency. *Frontiers in Human Neuroscience*, *7*, 373. https://doi.org/10.3389/fnhum.2013.00373

Ten Brink, A., Nijboer, T., Van Beekum, L., Van Dijk, J., Peeters, R., Post, M., & Visser, J. (2013). De Nederlandse Catherine Bergego schaal: Een bruikbaar en valide instrument in de CVA-zorg. *Wetenschappelijk Tijdschrift Voor Ergotherapie*, *6*(3), 27–36.

Ten Brink, A. F., Biesbroek, J. M., Oort, Q., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2019). Peripersonal and extrapersonal visuospatial neglect in different frames of reference: A brain lesion-symptom mapping study. *Behavioural Brain Research*, *356*, 504–515. https://doi.org/10.1016/j.bbr.2018.06.010

Ten Brink, A. F., Verwer, J. H., Biesbroek, J. M., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2017). Differences between left- and right-sided neglect revisited: A large cohort study across multiple domains. *Journal of Clinical and Experimental Neuropsychology*, *39*(7), 707–723. https://doi.org/10.1080/13803395.2016.1262333

Ten Brink, A. F., Visser-Meily, J. M. A., Schut, M. J., Kouwenhoven, M., Eijsackers, A. L. H., & Nijboer, T. C. W. (2017). Prism adaptation in rehabilitation? No additional effects of prism adaptation on neglect recovery in the subacute phase poststroke: A randomized controlled trial. *Neurorehabilitation and Neural Repair*, *31*(12), 1017–1028. https://doi.org/10.1177/1545968317744277

Tieri, G., Morone, G., Paolucci, S., & Iosa, M. (2018). Virtual reality in cognitive and motor rehabilitation: Facts, fiction and fallacies. *Expert Review of Medical Devices*, *15*(2), 107–117. https://doi.org/10.1080/17434440.2018.1425613

Tiku, M. L. (1968). Estimating the parameters of log-normal distribution from censored samples. *Journal of the American Statistical Association*, *63*(321), 134–140. https://doi.org/10.2307/2283834

Tinga, A. M., Visser-Meily, J. M. A., van der Smagt, M. J., van der Stigchel, S. V., van Ee, R., & Nijboer, T. C. W. (2015). Multisensory stimulation to improve low- and higher-level sensory deficits after stroke: A systematic review. *Neuropsychology Review*, *26*, 1–19. https://doi.org/10.1007/s11065-015-9301-1

Tobler-Ammann, B. C., Surer, E., Knols, R. H., Borghese, N. A., & de Bruin, E. D. (2017). User perspectives on exergames designed to explore the hemineglected space for stroke patients with visuospatial neglect: Usability study. *JMIR Serious Games*, *5*(3), e18. https://doi.org/10.2196/games.8013

Toyoda, K., Yoshimura, S., Nakai, M., Koga, M., Sasahara, Y., Sonoda, K., Kamiyama, K., Yazawa, Y., Kawada, S., Sasaki, M., Terasaki, T., Miwa, K., Koge, J., Ishigami, A., Wada, S., Iwanaga, Y., Miyamoto, Y., Minematsu, K., Kobayashi, S., & Japan Stroke Data Bank Investigators. (2022). Twenty-year change in severity and outcome of ischemic and hemorrhagic strokes. *JAMA Neurology*, *79*(1), 61–69. https://doi.org/10.1001/jamaneurol.2021.4346

Tuena, C., Pedroli, E., Trimarchi, P. D., Gallucci, A., Chiappini, M., Goulene, K., Gaggioli, A., Riva, G., Lattanzio, F., Giunco, F., & Stramba-Badiale, M. (2020). Usability issues of clinical and research applications of virtual reality in older people: A systematic review. *Frontiers in Human Neuroscience*, *14*, 93. https://doi.org/10.3389/fnhum.2020.00093

Turgut, N., Jansen, A.-L., Nielsen, J., Heber, I., Eling, P., & Hildebrandt, H. (2021). Repeated application of the covert shift of attention task improves endogenous but not exogenous attention in patients with unilateral visuospatial inattention. *Brain and Cognition*, *151*, 105732. https://doi.org/10.1016/j.bandc.2021.105732

Turner, S. L., Karahalios, A., Forbes, A. B., Taljaard, M., Grimshaw, J. M., Cheng, A. C., Bero, L., & McKenzie, J. E. (2020). Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: A review. *Journal of Clinical Epidemiology*, *122*, 1–11. https://doi.org/10.1016/j.jclinepi.2020.02.006

Van den Noortgate, W., & Onghena, P. (2003). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*(3), 325–346. https://doi.org/10.1521/scpq.18.3.325.22577

van der Kemp, J., Dorresteijn, M., Ten Brink, A. F., Nijboer, T. C. W., & Visser-Meily, J. M. A. (2017). Pharmacological treatment of visuospatial neglect: A systematic review. *Journal of Stroke and Cerebrovascular Diseases*, *26*(4), 686–700. https://doi.org/10.1016/j.jstrokecerebrovasdis.2017.02.012

van Ee, R., van Boxtel, J. J. A., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *The Journal of Neuroscience*, *29*(37), 11641–11649. https://doi.org/10.1523/JNEUROSCI.0873-09.2009

van Lieshout, E. C. C., van Hooijdonk, R. F., Dijkhuizen, R. M., Visser-Meily, J. M. A., & Nijboer, T. C. W. (2019). The effect of noninvasive brain stimulation on poststroke cognitive function: A systematic review. *Neurorehabilitation and Neural Repair*, *33*(5), 355–374. https://doi.org/10.1177/1545968319834900

Van Vleet, T. M., & DeGutis, J. M. (2013). The nonspatial side of spatial neglect and related approaches to treatment. *Progress in Brain Research*, *207*, 327–349. https://doi.org/10.1016/B978-0-444-63327-9.00012-6

Vleet, T. V., Bonato, P., Fabara, E., Dabit, S., Kim, S.-J., Chiu, C., Bisogno, A. L., Merzenich, M., Corbetta, M., & DeGutis, J. (2020). Alertness training improves spatial bias and functional ability in spatial neglect. *Annals of Neurology*, *88*, 747–758. https://doi.org/10.1002/ana.25862

Wafa, H. A., Wolfe, C. D. A., Bhalla, A., & Wang, Y. (2020). Long-term trends in death and dependence after ischaemic strokes: A retrospective cohort study using the South London Stroke Register (SLSR). *PLoS Medicine*, *17*(3), e1003048. https://doi.org/10.1371/journal.pmed.1003048

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: A review. *Frontiers in Psychology*, *10*, 158. https://doi.org/10.3389/fpsyg.2019.00158

Winstein, C., Kim, B., Kim, S., Martinez, C., & Schweighofer, N. (2019). Dosage matters. *Stroke*, *50*(7), 1831–1837. https://doi.org/10.1161/STROKEAHA.118.023603

How to cite this article: Huygelier, H., Tuts, N., Michiels, K., Note, E., Schillebeeckx, F., Tournoy, J., Vanden Abeele, V., van Ee, R., & Gillebert, C. R. (2024). The efficacy and feasibility of an immersive virtual reality game to train spatial attention orientation after stroke: A stage 2 report. *Journal of Neuropsychology*, *00*, 1–52. https://doi.org/10.1111/jnp.12403

APPENDIX A

DATA ANALYSIS (PLAN PRIOR TO START OF STUDY)

A.1 | DESCRIPTION OF PATIENT SAMPLE

We will report the mean, median, standard deviation, and range of a patient's age in years, years of formal education and time since stroke (i.e., days between hospital admission and first screening session). We will report the number of men and women. We will also report the results from the measure of anosognosia, post-stroke depression and post-stroke fatigue. We will report on which subtests patients performed lower than age-adjusted cut-offs on the OCS-NL and how patients performed on the neglect tasks at screening. In addition, we will report the number of patients with ischaemic versus haemorrhagic stroke and visualize the lesion locations. Lesions will be delineated on clinical computer tomography (CT) or Fluid-attenuated inversion recovery (FLAIR) or T2-weighted magnetic resonance imaging (MRI) scans using the Matlab Clusterize toolbox (de Haan et al., 2015) or manually using MRIcron. Scans will be converted from native to MNI space using age-specific CT and MRI templates of the Matlab SPM clinical toolbox (Rorden et al., 2012). Moreover, we will test whether there was a difference between the time since stroke between treatment groups A and B using a Bayesian paired *t*-test. These data will be reported for the per-protocol and intention-to-treat samples.

A.2 | QUALITY MONITORING OF CLINICAL TRIAL PROCEDURE AND MISSING DATA

We will report the number of protocol deviations, type of deviation and reason for deviations (e.g., technical problems, patients not available). In addition, to evaluate the extent to which patients and

assessors were blinded, we will report the percentage of accurate guesses of the intervention conditions. Furthermore, we will report the number of hours and types of daily therapy (e.g., occupational therapy, physical rehabilitation) that patients received in the rehabilitation unit. Last, we will report the ratio of contra- and ipsilesional stimulation in the IVR game and the total number of hours that patients used the IVR game in the placebo and active IVR conditions. Any technical problems with the IVR game will also be reported.

Our efficacy analyses will be performed on the *per-protocol sample* (i.e., patients who complete at least 80% of sessions), because it is crucial that there is no difference in the number of completed assessments between the placebo and active intervention for our within-subject comparison. To assess the extent to which the per-protocol sample represents the *intention-to-treat sample*, we will report on the prevalence of early drop-out and the missing data pattern. In addition, we will report the reasons for early drop-out.

A.3 | PLANNED ANALYSES

A.3.1. | What are the effects of an active IVR intervention compared to a placebo IVR intervention and its relationship with therapy dose on neglect symptoms outside the IVR environment?

Our *primary outcome measure* is the response times on invalid-cued targets for the shortest SOA on the Posner task. We will test whether there is a stronger decrease in the difference between left- and right-sided performance on the Posner task for the active compared to the placebo intervention. Moreover, we will test whether performance improves or remains stable for the left and right visual fields. In addition, we will evaluate the relation between therapy dose and the change in the left- and right-sided performance. To this end, we will estimate the primary outcome as a function of the main effects of therapy dose (i.e., 4, 8 and 10 h of intervention), type of intervention (i.e., placebo and active) and visual field (i.e., left and right side), their pairwise interactions and three-way interaction.

If there is no evidence in favour of a group treatment effect for our primary outcome, we will explore whether there is evidence in favour of between-subject differences in the treatment effect. To inform future research, we will then explore whether between-subject differences in the treatment effect may be related to anosognosia, post-stroke depression, post-stroke fatigue and lesion neuro-anatomy.

We will additionally evaluate the treatment effect on our *secondary outcomes*: the proportion of cancelled targets on the computerized cancellation task, the proportion of cancelled targets on the OCS-NL hearts cancellation task, CBS score and EWB score on the McIntosh line bisection task. For the secondary outcomes, we will test the main effect of the test moment (i.e., pre- or post-treatment), type of intervention and the effect of visual field (i.e., left versus right), their pairwise interactions and three-way interaction.

A.3.2. | Does non-spatial attention outside the IVR environment recover due to IVR spatial attention training?

Additionally, we will evaluate whether potential improvements as a result of the IVR training are specific to spatial attention or generalize to non-spatial attention. To this end, we will evaluate whether patients improve on the sustained attention to response task more during the active than the placebo intervention. To evaluate this hypothesis, we will estimate performance on the sustained attention to response task as a function of the test moment and intervention type and their interaction.

A.3.3. | Are training effects inside the IVR environment larger than outside the IVR environment?

If patients do not improve on any of our outcomes obtained outside the IVR environment, it is important to establish whether patients improved inside the IVR environment. The latter can clarify whether training effects in the IVR environment may not have transferred. To test this, we will compare the effect of our intervention between two within-subject conditions: the IVR assessment and the

computerized cancellation task. We selected these two outcome measures, as performance on these two tasks is measured on a similar scale (i.e., proportion of detected targets). To test our hypothesis, we will estimate the proportion of detected targets as a function of the main effect of test moment, intervention, visual field (i.e., left versus right) and task and include their interactions.

Additionally, to assess the extent to which patients improve inside the IVR environment, we will evaluate whether patients make more head movements towards the left visual field in the IVR assessment after the IVR training compared to before. We hypothesize a stronger increase in left-oriented head movements as a result of the active than the placebo IVR training. To test this hypothesis, we will estimate horizontal head orientation (i.e., negative values indicate left-sided and positive values right-sided orientations) as a function of the interaction of intervention condition and test moment.

A.3.4. | Is IVR rehabilitation feasible?

We will evaluate whether self-reported cybersickness and user experience predict the number of sessions completed by patients during the clinical trial. We will also compare cybersickness before and after IVR exposure and report descriptive statistics of the User Experience scale and of the user evaluations at the end of each game session. Moreover, we will document the number of referred and successfully recruited patients and reasons for excluding patients, following the Consolidated Standards of Reporting Trials guidelines (Moher et al., 2010). These feasibility results will be reported for the intention-to-treat sample.

A.4 | GENERAL MODELLING APPROACH

We will analyse the data with Bayesian mixed models using the R brms package (Bürkner, 2017), since mixed models can accurately model time-unstructured data (Andersen & Millen, 2013; Van den Noortgate & Onghena, 2003). Moreover, mixed models can clarify whether there were significant between-subject differences in treatment effects. To estimate the effect of the treatment on our secondary outcomes, we will use a multivariate regression model. The latter allows to statistically test for differences in effects between the secondary outcome measures. If the multivariate regression model does not fit well, we will use separate regression models for each outcome. We will use a lognormal or shifted lognormal distribution to estimate response times, which are typically skewed, and logistic regression to estimate proportion correct. We will evaluate model fit using posterior predictive checks (Gelman et al., 1996). We prefer a Bayesian approach because it allows to quantify the strength of evidence in favour of the null hypothesis (Wagenmakers, 2007).

In addition, we will calculate the Bayes Factor (BF) for contrasts using the paired *t*-tests of the Bayes Factor package (Rouder et al., 2009). We will interpret the BF according to the rule of Kass and Raftery (1995). A BF_{10} larger than 3 suggests substantial evidence, and larger than 10 suggests strong evidence in favour of the alternative model. BF_{10} smaller than 1/3 represents substantial, and smaller than 1/10 represents strong evidence for the null model. BF_{10} in between 1/3 and 3 represents inconclusive evidence.

A.5 | SAMPLE SIZE DETERMINATION

Efficient data collection in neglect rehabilitation research is important, as it is very difficult to recruit sufficient patients (Harvey, 2019). For this reason, we will use a sequential Bayes Factor design (Schönbrodt et al., 2015), in which we sequentially calculate the BF_{10} for our *main contrast of interest* to determine when to stop data collection. We predict a smaller difference between left- and right-sided response times to invalid-cued targets for the shortest stimulus onset asynchrony (SOA) in the Posner task (*R-L score*) after than before active treatment and we expect no difference in the R-L score after than before placebo treatment. Thus, our *main contrast of interest* is the pre-post active intervention difference in the R-L score compared to the pre-post placebo intervention difference in the R-L score. We will evaluate our main contrast of interest after 8, 16, 20 and 24 patients who completed at least 80% of the

TABLE A1 Characteristics and performance on the Posner task of 5 pilot participants.

	Patient characteristics				Response times (s) M (SD)				Proportion detected M (SD)			
Patient	Age (y)	Gender	TSS	Stroke side	L I	L V	R I	R V	L I	L V	R I	R V
P1[a]	62	M		RH	1.87 (1.28)	1.85 (1.34)	.61 (.17)	.59 (.36)	.20 (.41)	.33 (.47)	1.0 (.0)	1.0 (.0)
P2	48	M	48	RH	.44 (.03)	.43 (.06)	.39 (.06)	.38 (.03)	1.0 (.0)	1.0 (.0)	1.0 (.0)	1.0 (.0)
P3	58	F	91	RH	.92 (.48)	.77 (.37)	.76 (.38)	.62 (.17)	1.0 (.0)	1.0 (.0)	.98 (.16)	1.0 (.0)
P4	65	F	129	RH	1.18 (.67)	.81 (.39)	.73 (.47)	.54 (.14)	.98 (.16)	.95 (.22)	1.0 (.0)	1.0 (.0)
P5	44	M	113	RH	.70 (.33)	.66 (.25)	.52 (.15)	.59 (.24)	.80 (.41)	.73 (.45)	.76 (.42)	.95 (.22)

Abbreviations: L I, left invalid; L V, left valid; LH, left-hemispheric; M, mean; R I, right invalid; R V, right valid; RH, right-hemispheric; SD, standard deviation; TSS, time since stroke in days; y, years.

[a]The CT or MRI scan or radiologist report was not available to determine the date of the stroke. The hospital admission data was not available.

trial sessions. If the BF_{10} exceeds a threshold of 10 or .1 before reaching a sample size of 24 patients, we will stop data collection. Otherwise, we stop data collection when reaching a sample size of 24 patients.

To evaluate the probability of obtaining inconclusive or misleading evidence, we ran simulations following the principles of Schönbrodt and Wagenmakers (2018). First, we simulated response times on invalid-cued left and right targets for the 150 ms SOA on the Posner task. To obtain realistic estimates of the averages, between-subject, within-subject variances and associations between the variables, we based our simulated data on Posner data of 5 stroke patients (Table A1). We sampled effect sizes from a normal distribution (M = .2, SD = .10). The treatment effect was scaled for each case using the within-subject variance in response times. Thus, patients with a higher pre-treatment mean response time are expected to have a larger decrease in response times. This corresponded to a reduction of left-sided response times of 96 ms due to active treatment at the group level (Cohen's d = .28).

We estimated the BF_{10} using a Bayesian paired t-test of the Bayes Factor package (Rouder et al., 2009) for our main contrast of interest for 50,000 samples under the alternative hypothesis and 5000 samples under the null hypothesis, each consisting of 24 cases. These simulations revealed that 66% of trials under the alternative hypothesis resulted in a BF_{10} larger than 10 and 78.7% of trials resulted in a BF_{10} larger than 3. A total of 55% of trials reached a BF_{10} larger than 10 at 8 patients, .1% at 16, .02% at 20 and 11% at 24 patients. A total of 18.7% of trials resulted in inconclusive evidence at 24 participants. None of the trials resulted in strong evidence, and 2.6% of trials resulted in moderate-strength evidence in favour of the null hypothesis. If the null hypothesis was true, none of the BF_{10} exceeded the .10 threshold, while 1% of trials resulted in a BF_{10} exceeding 10 (i.e., a false positive result) before reaching 24 participants. A total of 65% of trials reached a BF_{10} smaller than 1/3 at 24 participants, 1.4% resulted in a BF_{10} larger than 3 (i.e., a false positive result) and 33% trials reached an inconclusive BF at 24 participants. We evaluated whether a sample size of 30 patients would produce better results, but a total of 27% of trials still reached an inconclusive BF at 30 patients.

Thus, the probability of a false negative result is 2.6%, and the probability of a false positive result is 2.4%. In the frequentist framework a power of 90% would correspond to a 10% probability of a false negative result. Thus, in comparison, our study has a low probability of producing false results.

APPENDIX B

DATA-ANALYSIS (ADJUSTED PLAN AFTER DATA COLLECTION)

As we did not reach the planned minimal per-protocol sample size for our group analysis, we adjusted our data-analysis plan accordingly.

B.1. | Training effects inside the iVR environment

To assess the extent to which patients improved inside the iVR environment, we compared performance on the iVR assessment between the baseline, the post-placebo and the post-active assessment. To this end, the proportion of detected targets was modelled as a function of the main effect of the location of the target relative to the head midline, the test moment and their pairwise interaction. A logistic regression model was estimated per case.

Additionally, we evaluated whether the head orientation changed as a function of training. The head orientation distributions were complex (characterized by skew, flat tops and multiple modes). Given the complexity of these distributions and the circular nature of the head orientation data (Cremers & Klugkist, 2018), we performed a non-parametric statistical test for circular data to compare the distributions from the R package TwoCircles per case (Jammalamadaka et al., 2021). We used the Rao test with Monte Carlo sampling and 5000 bootstrap samples. In addition, we described the median directions and mean resultant lengths (i.e., concentration) of the distributions using the R package circular (Lund et al., 2023). The mean resultant length is a parameter ranging from 0 to 1 representing how the circular

data are spread, with 0 representing the highest spread and 1 the lowest possible spread (Cremers & Klugkist, 2018).

B.2. | Training effects outside the iVR environment

Our main objective was to evaluate the efficacy of the iVR training on neglect symptoms outside the iVR environment. To this end we had several outcome measures of spatial bias and non-spatial attention (i.e., Sustained Attention to Response Test).

B.2.1. | Primary outcome measure

For the per-protocol cases, the primary outcome measure (response times on invalid cued targets) was modelled as a function of the active and the placebo therapy dosage (i.e., number of completed VR training trials) for both visual fields (left as the reference group) for each case separately:

$$y = \beta_0 + \beta_1 \text{ Right Target} + \beta_2 \text{ Placebo dosage} + \beta_3 \text{ Placebo dosage} * \text{Right Target} + \beta_4 \text{ Active dosage} + \beta_5 \text{ Active dosage} * \text{Right Target}$$

The regression coefficient β_4 reflects the change in response times resulting from the active therapy dosage. The therapy dosage was operationalized as the number of completed trials in the iVR training, as the number of trials was not matched between the active and placebo therapy conditions and because the number of trials reflects the main treatment manipulation. This variable was scaled by dividing it by 100 such that the regression coefficient represents the effect of completing 100 training trials.

In addition, for the intention-to-treat cases we estimated the following model:

$$y = \beta_0 + \beta_1 \text{ Right Target} + \beta_3 \text{ Therapy Dosage} * \text{Right Target}$$

The measurement moment at follow-up was not included in these models as we aim to establish the trend resulting from active therapy, rather than a mixed trend of active and no therapy.

B.2.2. | Secondary outcome measures

The secondary outcome measures were not registered at intermediate therapy dosages but did follow an interrupted time series design, often used in single-case research (Turner et al., 2020). The trend of change over phase 1 (placebo or baseline phase) and the change in slope relative to this trend were estimated for all cases using the following model:

$$y = \beta_0 + \beta_1 \text{ Right Target} + \beta_2 \text{ Days since start} + \beta_3 \text{ Days since start} * \text{Right Target} + \beta_4 \text{ Days since therapy} + \beta_5 \text{ Days since therapy} * \text{Right Target}$$

Thus, in this model, the regression coefficient (β_4) represents how the therapy impacted the trend in symptom change as established during the initial baseline phase (encompassing placebo treatment for per-protocol cases and treatment as usual for intention-to-treat cases). We predicted an accelerated improvement for left-sided targets as a result of the active therapy compared to the first phase (including baseline assessments for the intention-to-treat patients and placebo treatment for the per-protocol patients).

B.2.3. | Evidence synthesis

To integrate evidence across the secondary outcome measures of spatial bias, we used Bayesian Evidence Synthesis. This technique allows to synthesize evidence across multiple statistical tests (from different families). The technique developed by Kuiper et al. (2013) and integrated in the R BayesCombo package (Contrino & Lazic, 2017) was used. To this end, the regression coefficients and their posterior standard deviations were extracted from the regression models.

B.3. | Are training effects inside the IVR environment larger than outside the IVR environment?

To evaluate whether iVR training effects were more pronounced inside the iVR environment in comparison to outside the iVR environment, we compared the effect of our intervention between two within-subject conditions: the IVR assessment and the computerized cancellation task. We estimated the proportion detected targets as a function of the main effect of test moment (i.e., test moment immediately before active training and test moment directly after active training), visual field (i.e., left versus right) and task and included their interactions. This analysis was performed for all patients who completed active training at the single-case level.

B.4. | Model families

We analysed the data in R version 4.1.3 with all regression models estimated with the R brms package (Bürkner, 2017). For all models, patients were analysed separately, and the default priors were used. To estimate all models, we used a minimum of 2500 iterations and 4 chains. For models that did not converge with 2500 iterations, we increased the iterations to 5000 or 10,000. All models converged with R-hat values equal to 1. In addition, model convergence was checked by visually inspecting the posterior distributions and chains. No convergence issues were identified. To inspect model fit, we used graphical posterior predictive checks (Gelman et al., 1996). These checks did not reveal significant issues with the fit of the models. All analysis code, pre-processed data and results of the posterior checks can be found at 10.6084/m9.figshare.24873813.

The response times on the Posner cueing test were modelled with a right-censored shifted lognormal model. The shifted lognormal model is characterized by three parameters: mu (i.e., the mean of the lognormal distribution), sigma (i.e., the standard deviation of the lognormal distribution) and a shift (i.e., non-decision time; the minimal time required to make a response) and can model the skew typical for response times. We assumed that the mean and sigma of the lognormal distribution could change over repeated sessions but constrained the shift parameter to a constant value, as we did not expect changes in the speed of low-level sensorimotor processes over time. A censored model was chosen as patients often did not respond within the 4-s time limit for left-sided targets and because ignoring censoring results in biased estimates (Guo et al., 2022; Tiku, 1968). To analyse the cancellation tests, we used a logistic regression model. To analyse the McIntosh Line Bisection test and the scores on the Catherine Bergego Scale, we used a Gaussian regression model. Response times on the SART were modelled with a shifted lognormal model without censoring, as this provided a better fit than a censored model. False alarms on the SART were modelled using a logistic regression model.

APPENDIX C

DESCRIPTION OF LESION CHARACTERISTICS

We used the Lesion Quantification Toolkit (Griffis et al., 2021) to estimate the impact of the lesion on different regions of interest (parcels) included in the Yeo-Schaeffer 7-network atlas (Schaefer et al., 2018).

TABLE C1 Lesion loads (>25%) in different parcels of the Yeo-Schaefer atlas per case.

Hemisphere	Network label or main region	Parcel	X	Y	Z	Parcel loads (%)
Patient VR034						
R	Cont	Par 1	58	−38	44	31.2
R	Som Mot	4	58	−4	32	37.4
R	Default	PF Cv 1	35	28	−13	42.8
R	Dors Attn	Pr Cv 1	49	11	28	47
R	Default	PF Cv 2	51	29	1	57.7
R	Sal Vent Attn	Temp Occ Par 2	60	−25	29	73.7
R	Dors Attn	Post 2	49	−23	43	78.3
R	Corpus striatum	LN pallidum	28	6	3	86.2
R	Som Mot	1	53	−15	7	87.6
R	Sal Vent Attn	Fr Oper Ins 1	40	9	2	96.9
R	Som Mot	3	57	−3	13	97.8
R	Som Mot	2	41	−15	16	99.8
Patient VR060						
R	Vis	5	8	−75	6	23.7
R	Vis	6	17	−57	7	28.2
R	Cerebelum-Vermis	4 5	18	−42	−17	38.9
R	Cerebelum-Vermis	6	26	−57	−23	52.3
R	Vis	1	32	−30	−20	77.1
R	Vis	2	27	−64	−11	96
Patient VR076						
R	Default	Temp 1	61	−22	−17	26
R	Default	Temp 2	51	7	−16	27
R	Vis	2	27	−64	−11	30.7
R	Cont	P Cun 1	10	−64	43	33.5
R	Corpus striatum	LN pallidum	28	6	3	36.6
R	Cont	Cing 1	5	−26	34	43.3
R	Vis	4	22	−93	−3	46.4
R	Cont	PFCl 4	43	17	46	48.2
R	Sal Vent Attn	Med 2	8	7	54	50.8
R	Dors Attn	Post 5	14	−51	67	51.6
R	Som Mot	8	11	−23	66	56.5
R	Sal Vent Attn	Fr Oper Ins 1	40	9	2	57.1
R	Dors Attn	FEF 1	28	−1	60	58.4
R	Vis	3	49	−59	−10	73.4
R	Dors Attn	Post 4	27	−66	51	78.8
R	Cont	Par 2	45	−62	47	81.4
R	Som Mot	6	41	−21	61	81.9
R	Som Mot	7	30	−36	65	83.4
R	Dors Attn	Pr Cv 1	49	11	28	86.9
R	Cont	Par 1	58	−38	44	87.2

TABLE C1 (Continued)

Hemisphere	Network label or main region	Parcel	X	Y	Z	Parcel loads (%)
R	Som Mot	5	47	−9	49	91.3
R	Som Mot	4	58	−4	32	92.4
R	Dors Attn	Post 3	39	−44	50	94.8
R	Vis	7	36	−81	18	96.9
R	Default	Par 1	55	−50	32	97.7
R	Som Mot	3	57	−3	13	98
R	Sal Vent Attn	Temp Occ Par 2	60	−25	29	98.5
R	Sal Vent Attn	Med 1	11	−29	47	98.7
R	Default	Temp 3	57	−25	−1	98.8
R	Dors Attn	Post 1	51	−62	17	99
R	Som Mot	1	53	−15	7	99.1
R	Sal Vent Attn	Temp Occ Par 1	58	−41	14	99.4
R	Dors Attn	Post 2	49	−23	43	99.8
R	Som Mot	2	41	−15	16	100
Patient VR080						
L	Vis	8	−26	−87	22	28.4
R	Dors Attn	Post 3	39	−44	50	31.5
L	Cerebelum-Vermis	Crus 2	−28	−72	−37	32.3
L	Cerebelum-Vermis	Crus 1	−35	−66	−28	34.2
R	Corpus striatum	LN putamen	15	13	10	38.3
R	Dors Attn	Post 1	51	−62	17	44.9
R	Cont	PFCl 4	43	17	46	51.6
R	Limbic	Temp Pole 1	38	1	−34	52.2
L	Vis	5	−5	−91	−2	62.3
R	Default	Par 1	55	−50	32	62.9
L	Vis	4	−27	−93	−4	64.6
L	Vis	2	−25	−76	−13	72.8
R	Cont	PFCl 2	45	40	16	76.2
R	Dors Attn	Post 2	49	−23	43	80.5
R	Default	Temp 1	61	−22	−17	81
R	Cont	Par 1	58	−38	44	85.8
R	Som Mot	4	58	−4	32	88.2
R	Dors Attn	Pr Cv 1	49	11	28	90.5
R	Default	PF Cv 1	35	28	−13	92
R	Default	PF Cv 2	51	29	1	97.3
R	Default	Temp 3	57	−25	−1	97.4
R	Corpus striatum	LN pallidum	28	6	3	98.2
R	Thalamus	1	21	1	1	98.8
R	Som Mot	3	57	−3	13	99.5
R	Som Mot	1	53	−15	7	99.7
R	Sal Vent Attn	Temp Occ Par 1	58	−41	14	99.9

(Continues)

TABLE C1 (Continued)

Hemisphere	Network label or main region	Parcel	X	Y	Z	Parcel loads (%)
R	Default	Temp 2	51	7	−16	99.9
R	Sal Vent Attn	Temp Occ Par 2	60	−25	29	100
R	Sal Vent Attn	Fr Oper Ins 1	40	9	2	100
R	Som Mot	2	41	−15	16	100
Patient VR088						
R	Sal Vent Attn	Fr Oper Ins 1	40	9	2	29
R	Som Mot	1	53	−15	7	30.6
R	Som Mot	2	41	−15	16	53.3
R	Corpus striatum	LN pallidum	28	6	3	53.5
R	Thalamus	1	21	1	1	79.4
Patient VR108						
R	Thalamus	1	21	1	1	60
R	Corpus striatum	LN putamen	15	13	10	76.3
R	Corpus striatum	LN pallidum	28	6	3	83.1

Abbreviations: Cont, control network; Default, default mode network; Dors Attn, dorsal attention network; Sal Vent Attn, saliency/ventral attention network; Som Mot, somatosensory-motor network; Vis, visual network.

APPENDIX D

QUALITY MONITORING OF CLINICAL TRIAL PROCEDURE AND MISSING DATA

D.1 | PROTOCOL DEVIATIONS

Patient VR034 initially started participation in the clinical trial in October 2021. However, the patient developed a covid-19 infection during study participation and all visits needed to be cancelled. He re-enrolled in the study in April 2022. We only report the data from the second time he enrolled in the study. None of the other patients re-enrolled in our study.

All screening sessions, primary outcome assessments and VR game sessions were administered according to protocol. For the secondary outcome measures, missing data occurred for some cases. For patient VR034, the McIntosh Line Bisection Test was not administered in the third session due to time constraints. For patient VR060, all secondary outcome measures were administered as planned. Patient VR076 needed frequent breaks throughout sessions due to complaints of fatigue and low mood. For this reason, not all secondary outcome assessments were administered as planned (Table D1). Patient VR080 completed all assessments as planned, but the OCS-NL hearts cancellation test and computerized cancellation tests were not feasible due to low-level visual impairment (Table D1). Patient VR088 completed all assessments as planned. Last, for VR108, all secondary outcomes in the baseline assessment session were not administered due to an error of the examiner.

D.2 | BLINDNESS CHECK

Patient VR060 did not correctly guess the order of conditions he was assigned to, as he thought he was assigned to the Active-Placebo order of conditions. He indicated that the difference between the two conditions was not obvious. The blinded assessor also guessed that patient VR060 was assigned to the Active-Placebo order of conditions. In contrast, VR088 accurately guessed her order of conditions. The blinded assessor correctly guessed the order of conditions as well. Both were fairly confident in their choice.

TABLE D1 Overview of missing data and reasons for missingness on secondary outcome measures (1 = data complete).

Participant	Session	Cancellation	Hearts	McIntosh line bisection	SART	CBS
VR034	1	1	1	1	1	1
	2	1	1	1	1	1
	3	1	1	Time	1	1
VR060	1	1	1	1	1	1
	2	1	1	1	1	1
	3	1	1	1	1	1
	4	1	1	1	1	1
	5	1	1	1	1	1
VR076	1	1	1	1	1	1
	2	1	1	1	1	Fatigue
	3	1	1	1	Fatigue	Fatigue
VR080	1	Visual deficit	Visual deficit	1	1	1
	2	Visual deficit	1	1	1	1
	3	1	1	1	1	1
VR088	1	1	1	1	1	1
	2	1	1	1	1	1
	3	1	1	1	1	1
	4	1	1	1	1	1
	5	1	1	1	1	1
VR108	1	1	1	1	1	1
	2	Examiner error				
	3	1	1	1	1	1

Abbreviations: CBS, Catherine Bergego Scale; SART, sustained attention to response test.

APPENDIX E

REGRESSION MODELS: PARAMETER ESTIMATES

E.1. | Visualization of estimates

TABLE E1 Estimates of shifted lognormal model of response times on the Posner test as a function of treatment dosage and target location (A, active; P, placebo; R, right).

		Per-protocol patients						Intention-to-treat patients											
		VR060			VR088			VR034			VR076			VR080			VR108		
	Predictor	E	95% CI		E	95% CI		E	95% CI		E	95% CI		E	95% CI		E	95% CI	
Mu	Intercept	1.28	.86	1.74	−.34	−.64	−.06	−1.10	−1.25	−.94	1.65	1.26	2.14	.48	.20	.78	−1.22	−1.32	−1.11
	Treatment dosage (P)	**−.10**	**−.13**	**−.07**	−.02	−.05	.01	/	/	/	/	/	/	−.01	−.04	.03	−.01	−.02	.01
	Treatment dosage (A)	.01	−.01	.02	**−.02**	**−.03**	**−.01**	.01	−.00	.02	**−.08**	**−.14**	**−.02**	/	/	/	/	/	/
	Target location (R)	**−2.95**	**−3.44**	**−2.49**	**−1.11**	**−1.42**	**−.80**	**−.46**	**−.68**	**−.24**	**−1.83**	**−2.38**	**−1.37**	**−1.77**	**−2.08**	**−1.47**	**−.46**	**−.62**	**−.31**
	Treatment dosage (P)*Target location (R)	**.08**	**.04**	**.11**	.03	−.00	.06	/	/	/	/	/	/	−.00	−.04	.03	.02	.00	.04
	Treatment dosage (A)*Target location (R)	.00	−.01	.01	**.02**	**.01**	**.03**	−.01	−.03	.01	**.07**	**.01**	**.15**	/	/	/	/	/	/
Sigma	Intercept	.67	.49	.88	.48	.33	.64	−.30	−.45	−.15	.28	−.00	.60	.29	.13	.48	−.73	−.86	−.59
	Treatment dosage (P)	−.00	−.02	.01	**−.03**	**−.05**	**−.02**	/	/	/	/	/	/	−.00	−.03	.02	.01	−.01	.02
	Treatment dosage (A)	**−.01**	**−.02**	**−.00**	−.01	−.01	.00	−.00	−.01	.01	−.00	−.04	.04	/	/	/	/	/	/
	Target location (R)	**−.78**	**−1.03**	**−.54**	**−.98**	**−1.18**	**−.79**	.08	−.13	.29	−.26	−.64	.09	**−1.37**	**−1.59**	**−1.15**	**.27**	**.10**	**.47**
	Treatment dosage (P)*Target location (R)	**−.03**	**−.05**	**−.02**	.02	−.00	.04	/	/	/	/	/	/	.02	−.00	.05	/	/	/
	Treatment dosage (A)*Target location (R)	**.03**	**.02**	**.04**	.00	−.01	.01	.01	−.01	.03	−.01	−.06	.04	/	/	/	**−.04**	**−.06**	**−.02**
NDT	Intercept	.31	.30	.31	.27	.27	.27	.32	.31	.32	.10	.07	.11	.27	.25	.29	.25	.24	.26

Note: The left visual field is used as reference group. The estimates of the mu and sigma are in log units. The NDT parameter is in units of seconds.

Abbreviation: NDT, non-decision time (shift parameter). Bold values indicate 95% credible intervals that excluded zero.

TABLE E2 Estimates of logistic regression of proportion cancelled targets on the computerized and Hearts cancellation test as a function of days since the start of therapy and target location (A, active; P, placebo; R, right).

	Per-protocol patients						Intention-to-treat patients								
	VR060			VR088			VR034			VR076			VR080		
Predictor	**E**	**95% CI**		**E**	**95% CI**		**E**	**95% CI**		**E**	**95% CI**		**E**	**95% CI**	
Computerized cancellation test															
Intercept	−1.12	−1.40	−.86	−1.18	−1.46	−.93	.31	.07	.55	−3.58	−4.37	−2.92	/	/	/
Days since start	**−.07**	**−.09**	**−.06**	**.09**	**.07**	**.10**	−.01	−.02	.01	−.04	−.12	.03	/	/	/
Days since therapy	**.23**	**.18**	**.28**	**−.06**	**−.10**	**−.02**	.02	−.00	.05	.09	−.00	.18	/	/	/
Target location (R)	**1.92**	**1.57**	**2.29**	**.81**	**.45**	**1.18**	**.49**	**.14**	**.84**	**3.05**	**2.37**	**3.88**	/	/	/
Days since start*Target location (R)	**.04**	**.02**	**.06**	.01	−.01	.04	−.03	−.05	−.01	.03	−.04	.11	/	/	/
Days since therapy*Target location (R)	**−.12**	**−.18**	**−.07**	.04	−.06	.15	.04	.00	.08	−.01	−.11	.08	/	/	/
Hearts cancellation test															
Intercept	−1.69	−2.91	−.63	−.20	−.91	.52	.63	−.31	1.62	−2.40	−4.22	−1.02	/	/	/
Days since start	−.00	−.04	.04	**.07**	**.02**	**.12**	.03	−.04	.09	−.87	−3.21	−.03	/	/	/
Days since therapy	.06	−.02	.15	−.09	−.23	.03	−.08	−.19	.02	1.23	.03	4.61	/	/	/
Target location (R)	**1.50**	**.12**	**2.97**	.27	−.72	1.27	1.25	−.31	2.99	**2.21**	**.51**	**4.17**	/	/	/
Days since start*Target location (R)	.03	−.02	.08	−.03	−.09	.02	.04	−.09	.20	.89	.05	3.24	/	/	/
Days since therapy*Target location (R)	**−.17**	**−.29**	**−.06**	.08	−.08	.26	−.06	−.31	.14	−1.25	−4.64	−.04	/	/	/

Note: Estimates are in log odds. The left visual field is the reference condition. Bold values indicate 95% credible intervals that excluded zero.

TABLE E3 Estimates of linear regression of response position on the McIntosh Line Bisection test as a function of days since start of therapy and position of the right (R) and left (L) endpoint of the line and the CBS scores as a function of days since start and days since therapy (A, active; P, Placebo).

	Per-protocol patients						Intention-to-treat patients								
	VR060			VR088			VR034			VR076			VR080		
Predictor	E	95% CI		E	95% CI		E	95% CI		E	95% CI		E	95% CI	
McIntosh line bisection test															
Intercept	7.27	−1.94	16.47	−4.97	−9.51	−.66	/	/	/	18.04	12.63	23.50	7.32	.79	14.05
Days since start	**−.35**	**−.65**	**−.05**	−.01	−.20	.18	/	/	/	.05	−.31	.41	−.43	−1.06	.20
Days since therapy	**.92**	**.18**	**1.64**	−.05	−.53	.44	/	/	/	−.32	−.79	.14	.50	−.25	1.26
Left endpoint (8)	−4.17	−14.64	6.44	−1.05	−6.16	4.14	/	/	/	−4.15	−10.36	2.20	3.29	−4.58	10.97
Right endpoint (8)	**22.85**	**12.53**	**32.93**	**20.48**	**15.65**	**25.52**	/	/	/	**15.54**	**9.03**	**21.75**	**8.23**	**.56**	**16.31**
Days since start*Left endpoint (8)	−.25	−.60	.09	**−.24**	**−.47**	**−.02**	/	/	/	.14	−.30	.56	−.37	−1.08	.38
Days since start*Right endpoint (8)	−.32	−.65	.02	−.01	−.23	.21	/	/	/	.24	−.18	.68	.13	−.64	.84
Days since therapy*Left endpoint (8)	**.97**	**.15**	**1.83**	.45	−.15	1.04	/	/	/	−.17	−.71	.37	.34	−.57	1.17
Days since therapy*Right endpoint (8)	**.87**	**.04**	**1.69**	−.15	−.74	.42	/	/	/	.01	−.56	.55	−.12	−.94	.78
CBS															
Intercept	.43	−.06	.91	1.55	.90	2.21	.49	−.05	1.02	/	/	/	1.16	.57	1.75
Days since start	.01	−.03	.04	−.02	−.05	.01	−.01	−.05	.03	/	/	/	−.06	−.17	.04
Days since therapy	−.00	−.06	.06	.00	−.08	.08	.01	−.05	.06	/	/	/	.05	−.06	.17

Note: The estimates in the unit of millimetres for the line bisection test and the average score across items for the CBS. Bold values indicate 95% credible intervals that excluded zero.

Abbreviation: CBS, Catherine Bergego Scale.

TABLE E4 Estimates of logistic regression model comparing changes in task performance between VR and cancellation.

	Per-protocol patients						Intention-to-treat patients					
	VR060			VR088			VR034			VR076		
	E	95% CI		E	95% CI		E	95% CI		E	95% CI	
Intercept	−.72	−1.09	−.36	2.66	2.14	3.25	.72	.22	1.24	−2.25	−3.12	−1.53
Location: Right	1.86	1.34	2.41	2.21	.78	4.07	.21	−.53	.94	4.92	3.73	6.27
Time: Pre	−4.48	−6.45	−3.10	−.78	−1.47	−.10	−.54	−1.11	.01	−2.10	−3.48	−.82
Task: VR	1.69	1.15	2.23	−2.16	−2.85	−1.52	1.38	.61	2.20	2.53	1.71	3.49
Location * Time	2.65	1.18	4.66	−.94	−2.90	.73	−.32	−1.12	.49	−1.27	−2.93	.37
Location * Task	2.38	.34	5.50	−.75	−2.72	.85	1.99	.32	4.08	−3.90	−5.34	−2.58
Time * Task	2.92	1.43	4.94	−.63	−1.52	.27	.45	−.58	1.44	.78	−.60	2.22
Location * Time * Task	−2.40	−6.12	.62	1.82	−.12	3.99	−2.77	−5.01	−.86	.94	−.86	2.76

Note: The left visual field was the reference group. The estimates are in log odds.

TABLE E5 Estimates of logistic regression comparing performance on the VR assessment between baseline and post-training test moments.

	Per-protocol patients						Intention-to-treat patients											
	VR060			VR088			VR034			VR076			VR080			VR108		
Predictor	E	95% CI		E	95% CI		E	95% CI		E	95% CI		E	95% CI		E	95% CI	
Performance																		
Intercept	.60	.21	1.00	.13	−.13	.39	1.70	1.35	2.07	−.11	−.32	.09	−.43	−.79	−.09	3.15	2.53	3.86
Location	**.13**	**.10**	**.16**	.02	−.00	.03	−.02	−.04	.00	.03	.02	.05	**.09**	**.07**	**.12**	.00	−.04	.04
Post-active	**3.26**	**2.09**	**4.64**	**1.85**	**1.30**	**2.47**	1.86	.93	2.94	**1.14**	**.78**	**1.51**	/	/	/	/	/	/
Post-placebo	**1.75**	**.93**	**2.64**	.31	−.12	.73	/	/	/	/	/	/	**1.67**	**1.08**	**2.28**	1.62	−.02	4.07
Location*Post-active	.03	−.04	.09	**.07**	**.04**	**.10**	.11	.06	.17	−.00	−.02	.02	/	/	/	/	/	/
Location*Post-placebo	**.06**	**.00**	**.12**	**.07**	**.04**	**.09**	/	/	/	/	/	/	**.06**	**.02**	**.10**	.06	−.03	.17

Note: Bold values indicate 95% credible intervals that excluded zero.

TABLE E6 Estimates of regression model assessing changes over time in SART performance (response times and false alarms).

	Per-protocol patients						Intention-to-treat patients					
	VR060			VR088			VR034			VR080		
	E	95% CI		E	95% CI		E	95% CI		E	95% CI	
Response times (Go-Trials)												
Intercept	−.50	−.56	.45	−.89	.98	−.83	−.79	−.85	−.75	−.55	−.64	−.48
Days since start	−.00	−.00	.00	.00	−.00	.00	.01	.00	.01	−.02	−.02	−.01
Days since therapy	−.00	−.00	.00	.00	−.00	.01	−.00	−.01	.00	.02	.01	.02
Sigma	−.97	−1.08	−.86	−1.16	−1.26	−1.05	−1.22	−1.32	−1.11	−1.62	−1.75	−1.46
Sigma days since start	−.01	−.01	−.00	.00	.00	.01	−.01	−.02	−.00	−.00	−.01	.01
Sigma days since therapy	.00	−.01	.01	−.01	−.02	.00	.01	.00	.02	.00	−.01	.01
Non-decision time	.01	.00	.02	.02	.00	.05	.01	.00	.02	.04	.00	.08
False alarms (No-Go Trials)												
Intercept	−2.95	−4.75	−1.66	.00	−.62	.63	−.41	−1.22	.40	−2.13	−3.50	−1.01
Days since start	.04	.00	.08	−.01	−.04	.02	−.04	−.11	.02	.14	.04	.25
Days since therapy	−.08	−.17	.00	.02	−.05	.10	.04	−.05	.14	−.16	−.28	−.05

Note: Estimates for response times for mu and sigma are in log units and in seconds for the non-decision time. Estimates for false alarms are in log odds.

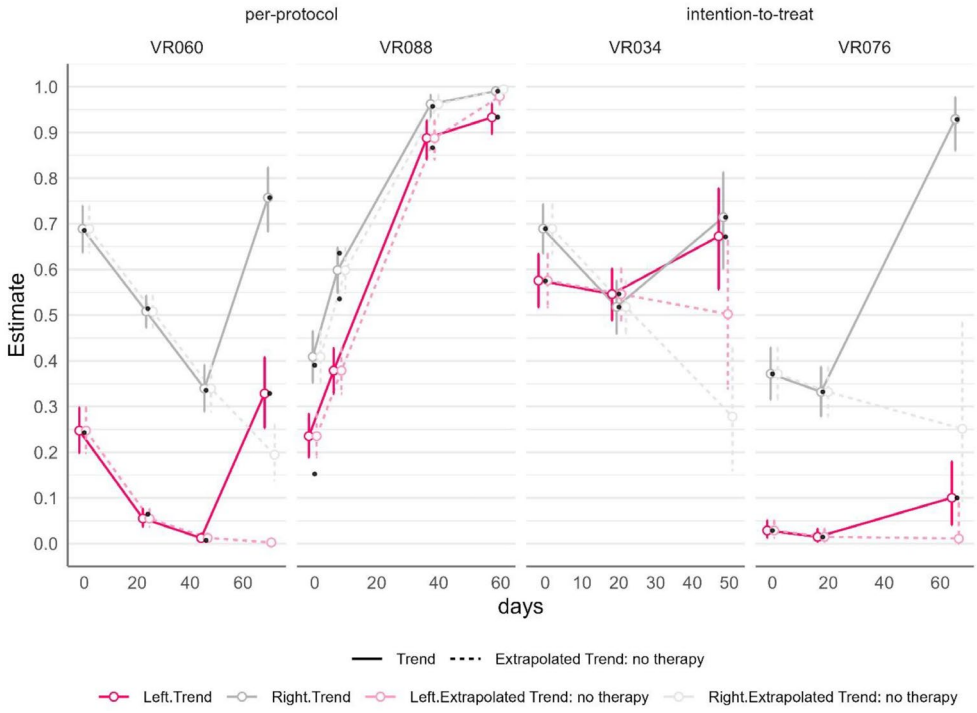

FIGURE E1 Estimated performance (probability of cancelling a target) on the computerized cancellation test as a function of days since the start of therapy for all cases for the left and right target locations. The solid line is the estimate of performance over days. The dashed line is the extrapolated trend established during the phase before active therapy started. Effects of active therapy are reflected in differences between the extrapolated trend and the observed trend. Error bars are 95% credible intervals of the posterior predictive distribution. The open dots are the estimates, and the solid dot is the observed value.

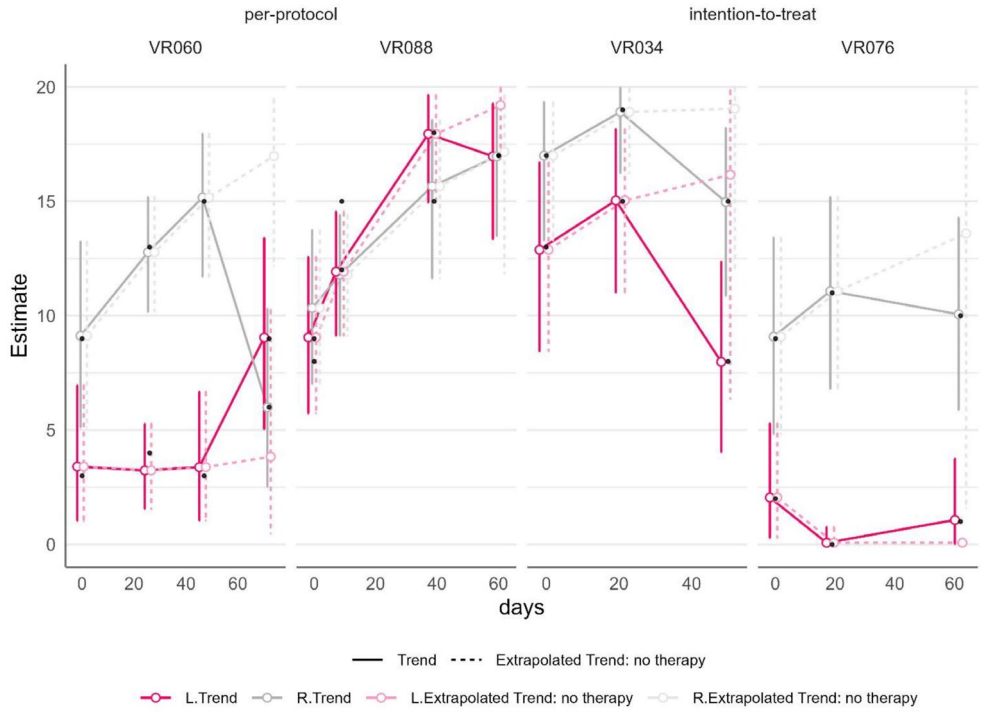

FIGURE E2 Estimated performance (total cancelled targets) on the Hearts cancellation test as a function of days since the start of therapy for all cases for the left and right target locations. The solid line represents the estimate of performance over days. The dashed line represents the extrapolated trend established during the pre-therapy phase. Effects of active therapy are reflected in differences between the extrapolated trend and the observed trend. Error bars are 95% credible intervals of the posterior predictive distribution. The open dots are the estimates, and the solid dot is the observed value.

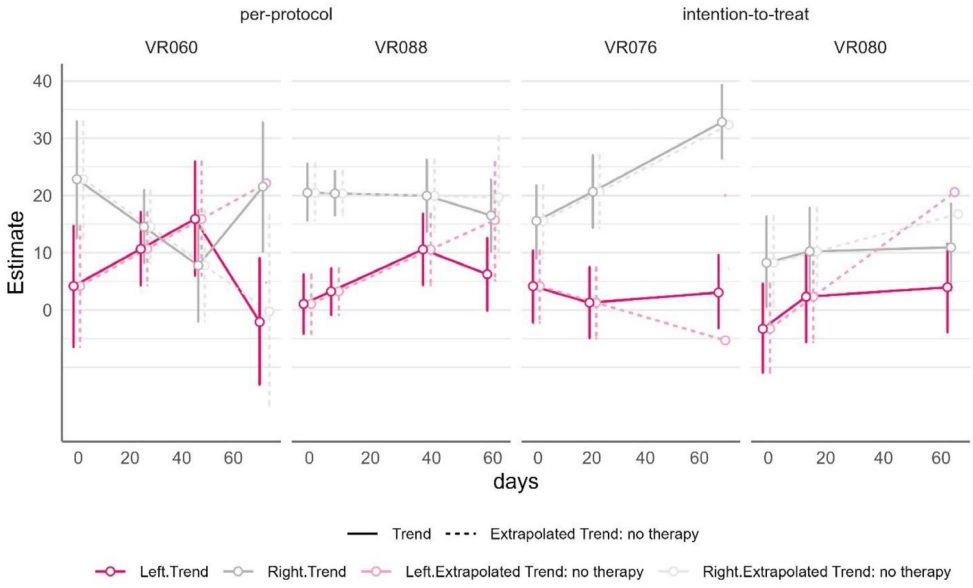

FIGURE E3 Estimated weights of the left and right endpoints on the McIntosh Line Bisection test as a function of days since the start of therapy for all cases. The solid line represents the estimates of performance over days. The dashed line represents the extrapolated trend established during the pre-therapy phase. Effects of active therapy (placebo therapy in VR080) are reflected in differences between the extrapolated trend and the observed trend. Error bars are 95% credible intervals of the posterior predictive distribution. The open dots are the estimates, and the solid dot is the observed value.

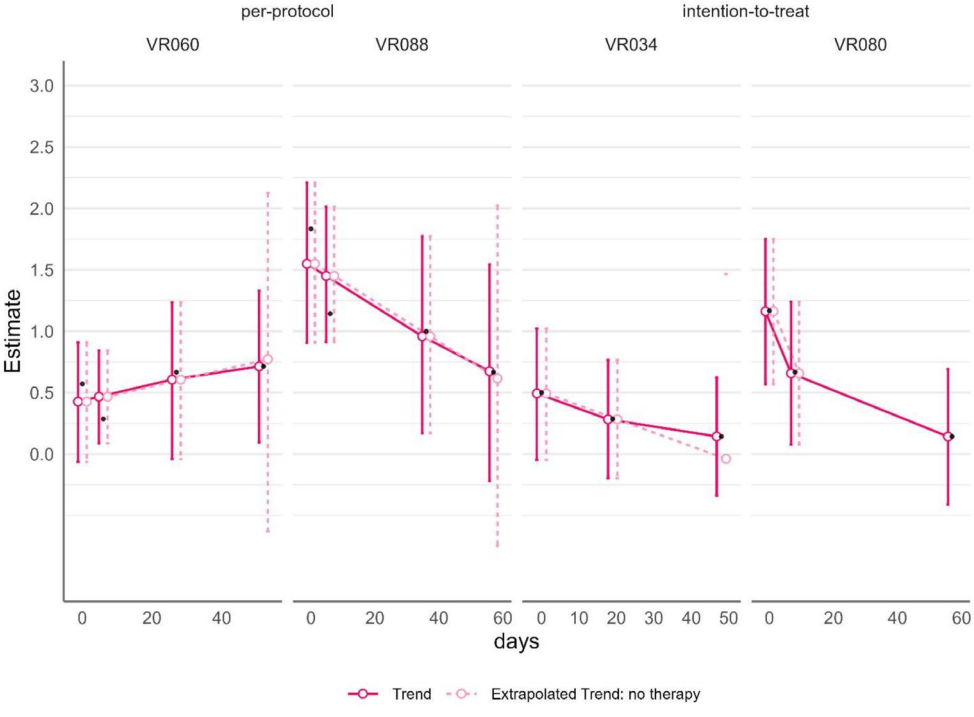

FIGURE E4 Estimated mean CBS scores as a function of days since the start of therapy for all cases. The solid line represents the estimates of performance over days. The dashed line represents the extrapolated trend established during the pre-therapy phase. Effects of active therapy (and placebo therapy for VR080) are reflected in differences between the extrapolated trend and the observed trend. Error bars are 95% credible intervals of the posterior predictive distribution. The open dots are the estimates, and the solid dot is the observed value.